WIRELESS ENGINEER

Vol. XXVII MARCH 1950 No. 318

Selectivity in Television

THE expansion of the British television service is creating some selectivity problems for designers. In view of its reputation for selectivity, therefore, it is not surprising that many are considering the possibility of the superheterodyne being more suitable than the t.r.f. set so widely used hitherto.

Most people, if asked which of the two types of receiver is the more selective, would unhesitatingly say the superheterodyne. A few would probably say that there is no difference between them. We shall show that the answer depends on circumstances and that the important factor is the relation of the product of bandwidth and circuit Q to signal frequency.

We are primarily interested in the television conditions and it is necessary to bear them carefully in mind. The main problem is to obtain sufficient selectivity in the vision channel to prevent the sound signal from interfering with the picture. It is, of course, one which has been present for years in receivers designed for the London station and for this a satisfactory solution has been found. The problem has been intensified in receivers designed for all other stations, however, by the adoption for them of vestigial-sideband transmission. This has made the solution appropriate to London inapplicable.

Television Standards

In all cases the sound-channel carrier lies 3.5 Mc/s below the vision-channel carrier, which carries modulation frequencies up to some 2.75 Mc/s. For the reception of London, where both sidebands are transmitted, there are three possibilities. Double-sideband reception can be used, in which case a vision-channel bandwidth of some 6 Mc/s at —6 db is needed and the sound-channel comes only 0.5 Mc/s below the lower edge of the vision pass-band. Alternatively, vestigial-sideband reception can be adopted with a 6-db bandwidth of only 3 Mc/s. The sidebands selected can be the upper or the lower. In the latter case, the sound channel again comes only 0.5 Mc/s below the vision pass-band, but if the upper sidebands are selected, the channel separation becomes 3.5 Mc/s.

Needless to say, this last case is a favourite with receiver designers and it is not difficult to obtain adequate discrimination between the two channels. The natural selectivity of the vision-channel intervalve couplings, supplemented by one rejector circuit tuned to the sound signal, usually suffices.

In the case of the new Midland station, and in all future stations, vestigial-sideband transmission is adopted and it is the lower-frequency sidebands, lying between the vision and sound carriers, which are employed. The receiver designer now has no choice but to adopt a 3-Mc/s vision-channel bandwidth with a 0.5-Mc/s separation from the sound channel.

In addition, he has to look forward to the time when more stations are operating and when the sound channel of another station may be 1.5 Mc/s above the vision carrier. Although such a signal may be weak in many districts it will not always be so. Moreover, it will be capable of producing more severe interference than the sound signal belonging to the wanted station. The direct beat between the carriers of the sound and vision signals of one station is at 3.5 Mc/s. Even if this frequency is transmitted through the vision channel, the resulting pattern on the picture is so

www.americanradiohistory.com

fine that it is not very noticeable. The important interference comes from the modulation on the sound channel.

With the sound signal of another station 1.5 Mc/s above the vision carrier, however, the beat will be 1.5 Mc/s and it will cause a very noticeable pattern on the screen. Because of this, it has been recommended that 50-db attenuation should be provided against the sound signal of an adjacent station, while only 30 db is needed against the sound signal of the wanted station. As a result, the receiver must have a vision-channel response curve somewhat as sketched in Fig. 1. The frequency scale has signs corresponding to the case of an i.f. amplifier in a superheterodyne having the local oscillator frequency above the signal frequencies. If the oscillator frequency is below the signal and for a t.r.f. set, the signs must be reversed.

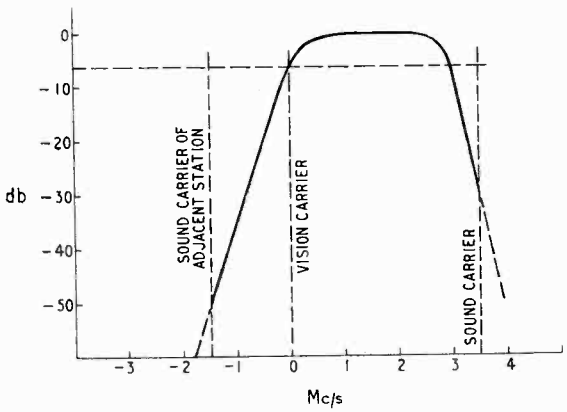


Fig. 1

After this preliminary statement of the selectivity requirements for television, we can turn to our main theme—the relative merits of the superheterodyne and the straight set. In the first place, there is nothing at all in the frequency-changing process which affects selectivity. All the frequency differences existing before the process persist unchanged after it. An interfering signal Δf away from the carrier of the wanted signal is still Δf away from it no matter to what intermediate frequency the signal is changed. Any difference in selectivity which may exist between superheterodyne and straight set is thus due solely to the difference of operating frequency.

The characteristics of a parallel-tuned circuit, such as that of Fig. 2, are most simply expressed as

$$S = R/Z = \sqrt{[1 + Q^2(1 - x^2)^2/x^2]} \quad .. \quad (1)$$

where $Q^2 = CR^2/L$, $x = \omega\sqrt{LC} = \omega/\omega_0 = f/f_0$, $f_0 = 1/(2\pi\sqrt{LC})$, and S is the ratio of the resonant impedance to the impedance away from resonance.

If we write $f = f_0 + \Delta f$ and $\Delta f \ll f_0$ we get the approximate relation

$$S^2 \approx 1 + 4Q^2\Delta f^2/f_0^2 \quad .. \quad .. \quad .. \quad (2)$$

An alternative, and often more useful, result is obtainable. There are two frequencies $f_1 < f_0$ and $f_2 > f_0$ for which the values of S are equal. If $n = f_2 - f_1$

$$S^2 = 1 + Q^2n^2/f_0^2 \quad .. \quad .. \quad .. \quad (3)$$

exactly. If f_0 were $(f_1 + f_2)/2$, n would be $2\Delta f$ and equations (2) and (3) would be the same, However, $f_0 = \sqrt{(f_1f_2)}$ and there is some difference between them. Now $Q^2 = CR^2/L$ and $f_0^2 = 1/4\pi^2LC$; therefore $Q/f_0 = 2\pi CR$. The equations can thus be written

$$S^2 \approx 1 + 16\pi^2C^2R^2\Delta f^2 \quad .. \quad .. \quad .. \quad (2a)$$

$$S^2 = 1 + 4\pi^2C^2R^2n^2 \quad .. \quad .. \quad .. \quad (3a)$$

At any level of response S the bandwidth $n = f_2 - f_1$ is quite independent of the resonance frequency and dependent only on the CR product. The same thing is true of Δf as long as $\Delta f \ll f_0$. When it is not, the resonance curve is, in effect, tilted slightly and for equal values of Δf the response is greater on one side of resonance than it is on the other. To put it differently $f_2 - f_0 > f_0 - f_1$.

Although we have examined here only a single resonant circuit the results with combinations of circuits are much the same. Cascades of single circuits all resonant to the same frequency obviously are equally independent of the resonance frequency. Stagger-tuned single circuits in cascade are also independent of the mid-band frequency, although this is less obvious, and coupled circuits have similar characteristics. What is true of the single circuit can be taken as being generally true of more complex networks, although there may be minor discrepancies in particular cases when $\Delta f/f_0$ is large.

For the reception of a given signal a certain minimum bandwidth n_1 is needed and it is often specified for an edge response of —6 db, for which $S^2 = S_1^2 = 4$. For selectivity some other bandwidth n_2 is required for which the response is S_2. From Equ. (3a) we have

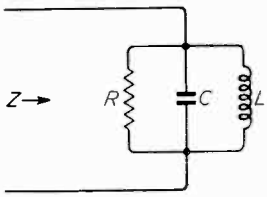


Fig. 2

$$n_2 = \frac{\sqrt{S_2^2 - 1}}{2\pi CR} \text{ and } n_1 = \frac{\sqrt{S_1^2 - 1}}{2\pi CR}$$

hence the difference of the bandwidths, $n_2 - n_1$ is constant and quite independent of f_0. With any given form of amplifier no control over the selectivity is obtainable by any change of f_0 so long as

www.americanradiohistory.com

the pass-band is not greater than that needed for the signal.

In view of this, how did the tradition of the superior selectivity of the superheterodyne arise and why is it still normally used in cases where selectivity is all important?

The answer is that in many common circumstances it is impracticable to make the pass-band only just wide enough for the transmission of the desired signals when the circuits are operating at the frequency of that signal. Equ. (3a) may be misleading here, for it gives the impression that the bandwidth can be reduced to any desired extent merely by increasing C but, in practice, R is likely to vary with an alteration of C. Equ. (3) gives a truer picture and at a given frequency bandwidth depends only on Q. At any frequency it is proportional to f_0/Q, but there is no simple and precise relation between f_0 and Q. It is found, however, that for coils of a given physical size and type Q is roughly proportional to $\sqrt{f_0}$. The bandwidth is thus roughly proportional to $\sqrt{f_0}$. Consequently for a given bandwidth requirement, there is always some frequency above which it is impossible, impracticable or uneconomical to obtain circuits of high enough Q to meet the requirement.

Receiver Type

Comparing t.r.f. and superheterodyne receivers having the same form and number of circuits affecting selectivity, the use of the latter is advantageous only when it is impracticable to reduce the bandwidth of the t.r.f. set to the minimum necessary for passing the signal. For the reception of sound signals, and generally whenever n/f_0 is very small, the superheterodyne is advantageous. Even with the bandwidth of 2—20 Mc/s needed in radar f_0 is so high that the superheterodyne would probably have to be used on selectivity grounds even if it were not already necessary for amplification.

Even in cases where adequate selectivity is possible at signal frequency, the use of the superheterodyne may still be necessary. When a receiver must be tunable over a band for station selection, there is a limit to the number of circuits which can be economically provided at signal frequency. This limit is usually around two or three only. Selectivity may demand the use of a dozen or more circuits, and the proper course is obviously to put most of them at the fixed frequency of the i.f. amplifier.

In the television case, it is easy to secure the minimum bandwidth in the vision channel at signal frequency and up to the present there has been no call for tunable receivers. On the sound side the minimum bandwidth for the transmission of the signal cannot be obtained at signal frequency, but this is not important for, in spite of this, adequate selectivity can still be secured.

It would appear, therefore, that in television the superheterodyne would offer no advantage over the t.r.f. set. This is quite true so long as the receiver does not include rejector circuits for sound-channel rejection or when such rejectors can have a considerable bandwidth. This condition exists in vestigial-sideband reception of the upper sidebands of the London transmitter and then there is little or nothing to choose between the two forms of receiver.

For other stations, however, vestigial-sideband reception of the lower sidebands is necessary. Rejectors of narrow bandwidth are then needed to prevent interference from the sound channel while leaving the vision-channel bandwidth unimpaired. If the required rejector bandwidth is less than can be obtained at signal frequency the superheterodyne has the advantage.

The choice between the two thus hinges not at all upon the main characteristics of the vision channel but entirely upon whether or not it is feasible to obtain rejector circuits of high-enough Q at signal frequency. In connection with rejector bandwidth it should be pointed out that, if it is attainable, it is possible to use a narrower bandwidth in the t.r.f. set than in the superheterodyne because in the latter some allowance must be made for frequency drift of the local oscillator.

We should not like to have to forecast what the eventual decision of designers will be, assuming no extraneous matters enter, for, although it is very hard to obtain a high-enough Q for proper rejector operation at frequencies around 60 Mc/s there are alternatives to the simple rejector. Some bridged-T networks, for instance, enable the equivalent of a rejector with infinite Q to be obtained. The decision will probably be made on extraneous grounds, however, such as the need for a tunable receiver and if so, it will almost certainly be in favour of the superheterodyne.

W. T. C.

www.americanradiohistory.com

SYNTHESIS OF WIDEBAND TWO-PHASE NETWORKS

By H. J. Orchard, B.Sc.

SUMMARY.—Hitherto, networks providing a two-phase supply from a single-phase supply over a wide band of frequencies have been designed empirically. A synthesis technique is now available which gives exact design formulae both for the network components and also for the relations between the design parameters. It is shown that the most general circuit meeting the requirement is essentially a pair of all-pass networks. By utilizing elliptic functions, such networks can be designed to have a Tchebichef approximation to the ideal requirement: this represents the most efficient condition. Simple computing schemes, design curves and a numerical example are included.

1. Introduction

IN the design of single-sideband modulator circuits and various other pieces of equipment, pairs of transmission networks are called for having the property that the two members of a pair shall give equal attenuations and have phase shifts differing from one another by $\pi/2$ radians over a wide band of frequencies.

Several publications [1, 2, 3] have been made concerning the theory of such network pairs and various design methods have been proposed, some graphical and some analytical. However, they all proceed, essentially, by analysing some special configuration and trying to adjust its parameters empirically to meet the requirements. This technique is time consuming and, moreover, the methods give no indication of the best performance of which the networks are capable for a given complexity.

Recently it has been found possible to place the whole of the theory on a completely analytical basis which yields both exact design formulae and also clear-cut information as to the theoretical limitations of any possible networks which might be capable of meeting the requirements.

2. General Theory

The precise requirement on the network pair is illustrated in Fig. 1. From one harmonic voltage, V_0 two more harmonic voltages, V_1 and V_2 must be derived, such that V_1 and V_2 are equal in magnitude but different in phase from one another by $\pi/2$ radians. The system may be regarded as a device for producing a two-phase supply from a single-phase supply. However, it is required to operate, not merely at one frequency, but equally well over a specified band of frequencies possibly many octaves wide. It is in this aspect of the problem that all the difficulty lies.

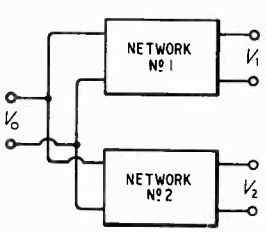


Fig. 1. *Basic phase-splitting circuit.*

Now the phase shift which occurs in any network may, following the nomenclature introduced by Bode[4], be considered as being composed of the minimum phase shift which can accompany the attenuation of the network, and which is uniquely defined by it, together with the phase shift of an all-pass network or phase-equalizer network which may be 'concealed' in the network. If this latter part is zero the network is a minimum-phase network. Thus:

(Total phase shift) = (Minimum phase shift) + (all-pass phase shift).

But both network No. 1 and network No. 2 in Fig. 1 must have equal attenuations at all frequencies in order that V_1 and V_2 shall be equal in magnitude. Hence the minimum phase shifts for the two networks must be identical. Therefore if V_1 and V_2 are to be equal for all frequencies they can differ in phase only by an amount which is the difference in phase shift between two all-pass networks. It is pointless to consider, consequently, the case where the two networks, No. 1 and No. 2, are other than all-pass networks, for it would imply that superfluous components existed in the networks, producing identical extra phase shift.

The problem thus resolves itself into one of finding two all-pass networks which have phase-shifts differing from one another by $\pi/2$ radians over a prescribed band of frequencies.

3. Properties of All-pass Networks

The most general all-pass network is a symmetrical lattice whose arms are inverse reactances; i.e., the product of their impedances is equal to the square of a constant resistance. Such networks are characterized by the fact that they have a resistive characteristic impedance, independent of frequency and zero attenuation for all real frequencies. However, at real frequencies they do introduce a phase-shift and, within some restrictions, this may be given a

MS accepted by the Editor, August 1949

www.americanradiohistory.com

wide variety of shapes by altering the arm reactances in the lattice. One restriction is that the slope of the phase-shift v. frequency curve must be positive.

The simplest all-pass network is shown in Fig. 2(a) and the one next in complexity in Fig. 2(b): these are referred to as first-degree and second-degree networks respectively. Every all-pass network of higher degree can be replaced by a tandem connection of suitable first- and second-degree networks and such a tandem connection is, from the practical point of view, usually preferable.

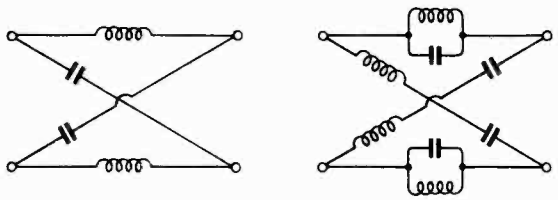


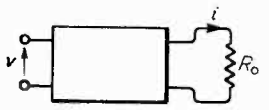
Fig. 2. *All-pass networks.*

The behaviour of an all-pass network, whatever its complexity, is most simply described in terms of its transfer impedance as defined in Fig. 3. If R_0 be the characteristic impedance of the network and Z_T the transfer impedance,

then $Z_T = v/i$

For design purposes it is most convenient to normalize the impedances of the network, taking $R_0 = 1$: this is finally changed to some suitable value at the end of the design. The transfer impedance is a rational function, with real coefficients, of the variable $p = j\omega$. The zeros of Z_T, which are identical with the transient modes of the terminated network, must, from considerations of stability, lie in the left half of the complex p-plane. In order that the impedance shall be appropriate to an all-pass network its modulus must be unity for all real frequencies,

Fig. 3. *Circuit for defining transfer impedance.*



and this requirement forces the poles of Z_T to be equal in number to, and the negative of, the zeros. Hence all the poles lie in the right half of the complex p-plane and the transfer impedance may be written as

$$Z_T = \frac{A + pB}{A - pB} \quad .. \quad .. \quad .. \quad .. \quad (1)$$

where A and B are even polynominals in p. It will be seen that, at real frequencies, the denominator polynomial in Equ.(1) is the complex conjugate of the numerator, and this forces the requirement of unit modulus on to Z_T.

The pole-zero patterns in the complex-plane appropriate to first- and second-degree networks are shown in Fig. 4(a) and 4(b) respectively, where circles denote zeros and crosses, poles. Due to the reality of the coefficients the poles and zeros must, if complex, occur in conjugate pairs.

The phase shift introduced by the all-pass network at real frequencies is equal to the phase angle of Z_T and consequently is given by

$$\tan \tfrac{1}{2}\beta = \frac{\omega \,.\, B(-\omega^2)}{A(-\omega^2)} \quad .. \quad .. \quad .. \quad (2)$$

where β is the phase shift in radians. Numerator and denominator polynomials introduce equal contributions to the phase at real frequencies as they are conjugates and so it is necessary to consider only the numerator.

Having thus summarized the properties of all-pass networks and their transfer impedances-from such aspects as are necessary,* the next problem is to see how two transfer impedances can be derived, suitable for all-pass networks, and having $\pi/2$ radians difference between their phase angles.

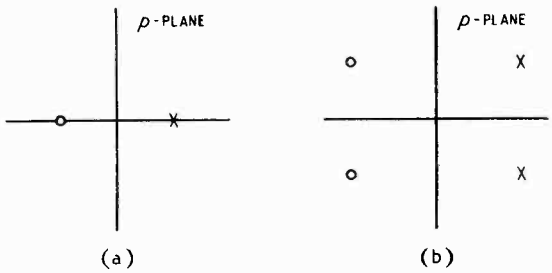


Fig. 4. *Pole-zero patterns.*

4. Derivation of Transfer Impedances

Let network No. 1 have a transfer impedance

$$Z_{T1} = \frac{A_1 + pB_1}{A_1 - pB_1} \quad .. \quad .. \quad .. \quad (3)$$

and network No. 2 have

$$Z_{T2} = \frac{A_2 + pB_2}{A_2 - pB_2} \quad .. \quad .. \quad .. \quad (4)$$

We have

$$\tan \tfrac{1}{2}\beta_1 = \frac{\omega B_1}{A_1}\,;\ \tan \tfrac{1}{2}\beta_2 = \frac{\omega B_2}{A_2} \quad .. \quad (5)$$

and hence, by the addition theorem,

$$\tan \tfrac{1}{2}(\beta_1 - \beta_2) = \frac{\dfrac{\omega B_1}{A_1} - \dfrac{\omega B_2}{A_2}}{1 + \dfrac{\omega^2 B_1 B_2}{A_1 A_2}}$$

$$= \frac{\omega(A_2 B_1 - A_1 B_2)}{(A_1 A_2 + \omega^2 . B_1 B_2)} \quad .. \quad .. \quad (6)$$

* For further properties see Ref. (4) Chap. 11.

www.americanradiohistory.com

$$\left.\begin{aligned} &\text{Put } P = A_2B_1 - A_1B_2 \\ &\text{and } Q = A_1A_2 + \omega^2 B_1 B_2 \end{aligned}\right\} \quad \ldots \quad \ldots \quad (7)$$

so that

$$\tan \tfrac{1}{2}(\beta_1 - \beta_2) = \frac{\omega P}{Q} \quad \ldots \quad \ldots \quad (8)$$

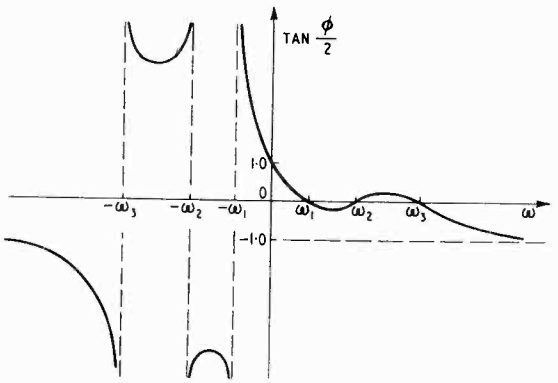


Fig. 5. *Typical curve for tan $\phi/2$.*

It will be seen that P and Q are even polynomials in $p = j\omega$. Over the prescribed band of frequencies we require $(\beta_1 - \beta_2)$ to approximate, as accurately as possible, to $\pi/2$ and so we are interested, not so much in the magnitude of $(\beta_1 - \beta_2)$, as in the error which is involved; i.e., in the quantity $\left\{\frac{\pi}{2} - (\beta_1 - \beta_2)\right\}$. Forming the tangent of half this error we get

$$\tan\left\{\frac{\pi}{4} - \frac{1}{2}(\beta_1 - \beta_2)\right\} = \frac{1 - \dfrac{\omega P}{Q}}{1 + \dfrac{\omega P}{Q}}$$

$$\text{or } \tan\frac{\phi}{2} = \frac{Q - \omega P}{Q + \omega P} \quad \ldots \quad \ldots \quad \ldots \quad (9)$$

where $\phi = \frac{\pi}{2} - (\beta_1 - \beta_2)$.

Thus $\tan\frac{\phi}{2}$ is seen to be a rational function whose poles and zeros are the negatives of one another, for the denominator polynomial is derived from the numerator by reversing the sign of ω (or p). Also for $\omega = 0$, $\tan\frac{\phi}{2} = 1$. Therefore we can write

$$\tan\frac{\phi}{2} = \prod_{\sigma=1}^{n}\left(\frac{\omega_\sigma - \omega}{\omega_\sigma + \omega}\right) \quad \ldots \quad \ldots \quad \ldots \quad (10)$$

where the ω_σ terms are the zeros of the polynomial $(Q - \omega P)$. As long as the ω_σ terms are either real or occur in conjugate pairs they are quite arbitrary. For the purpose of the present problem it is almost obvious that to get a good approximation over the prescribed band they should all occur at real values of ω and be distributed over that band. Then the typical form for the function $\tan\frac{\phi}{2}$ is as illustrated in Fig. 5.

The problem is now clearly one of finding the best approximation of this rational function over the prescribed band and, under these conditions, of obtaining a numerical relationship between bandwidth, error and the number of ω_σ terms required; i.e., the complexity of the networks.

The most efficient form of approximation is when the ω_σ terms are so chosen that the maximum excursions of $\tan\frac{\phi}{2}$ (and hence of ϕ) from zero are all equal and arranged as shown in Fig. 6. This is the so-called Tchebichef form of approximation and the problem of producing it with a rational function usually involves elliptic functions.

The method of obtaining the Tchebichef approximation is to represent the rational function and its variable by a pair of parametric equations in a manner analogous to that by which, for example, the cycloid is represented by the two equations.

$$y = a(1 - \cos t) \quad x = a(t - \sin t)$$

where the parameter is t. For the rational function, however, the parametric equations require not circular but elliptic functions.

Before describing in detail the particular

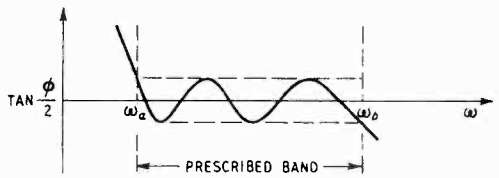


Fig. 6. *Tchebichef approximation in tan $\phi/2$.*

parametric equations needed here for $\tan\frac{\phi}{2}$ and ω it may be of interest to recall some of the basic properties of the elliptic functions. The elliptic functions are generalizations of the elementary circular functions sin, cos, etc., and stand midway, so to speak, between them and the hyperbolic functions. The most common one is the sn function which is defined by the integral

$$u = \int_0^{\operatorname{sn}(u)} \frac{dt}{\sqrt{(1 - t^2)(1 - k^2t^2)}}$$

the other elliptic functions are defined by similar integrals.

The integral contains the constant k which is called the modulus. (This is an unfortunate

www.americanradiohistory.com

choice of name and it should not be confused with its alternative meaning ' absolute magnitude of.') The elliptic functions are thus seen to be functions of two variables, the main variable u, and the modulus k. This is denoted symbolicaly in the sn function for example by writing it as sn $(u\ ;\ k)$, and similarly for the others. In general, k may be any complex number, but in most practical applications, as here, it is restricted to be real and lie between 0 and 1.

When $k = 0$ the functions degenerate into circular functions, and when $k = 1$, into hyperbolic functions. For these two special cases the sn integral becomes :

$$u = \int_0^{\sin(u)} \frac{dt}{\sqrt{(1 - t^2)}} \text{ and}$$

$$u = \int_0^{\tanh(u)} \frac{dt}{(1 - t^2)} \text{ respectively.}$$

Now the fundamental property of the circular and hyperbolic functions is that they are periodic : the former have a single real period of π or 2π ; the latter have a single imaginary period of $j\pi$ or $j2\pi$. The elliptic functions combine these two properties and have both a real and an imaginary period.

The real quarter-periods of the circular and elliptic functions are given by

$$\frac{\pi}{2} = \int_0^1 \frac{dt}{\sqrt{(1 - t^2)}} \text{ and}$$

$$K = \int_0^1 \frac{dt}{\sqrt{(1 - t^2)(1 - k^2t^2)}} \text{ respectively,}$$

while the imaginary quarter-periods of the hyperbolic and elliptic functions are given by

$$j\frac{\pi}{2} = \int_0^{j\infty} \frac{dt}{(1 - t^2)} \text{ and}$$

$$jK' = \int_0^{j\infty} \frac{dt}{\sqrt{(1 - t^2)\ (1 - k^2t^2)}} \text{ respectively.}$$

or, with a slight change of the variable of integration, $t = j\dfrac{x}{\sqrt{1 - x^2}}$

$$\frac{\pi}{2} = \int_0^1 \frac{dx}{\sqrt{(1 - x^2)}} \text{ and}$$

$$K' = \int_0^1 \frac{dx}{\sqrt{(1 - x^2)(1 - k'^2x^2)}}$$

$k' = \sqrt{1 - k^2}$ is called the complementary modulus.

Due to the double periodicity of the elliptic functions, the complex plane of its variable u is divided up into an infinte lattice of rectangles which are such that, at geometrically similar points in all the rectangles, the function has the same value ; these rectangles are called the period rectangles. The sides of the rectangles are $4K$ (or $2K$) and $j4K'$ (or $j2K'$), depending upon the particular function.

Another way of looking at this property is to consider that the complete complex plane of the elliptic function is mapped conformally on to each one of the period rectangles in the complex plane of its variable. This happens also with the circular and hyperbolic functions except that the rectangles degenerate into infinite strips parallel to either the imaginary axis or the real axis respectively.

When the elliptic functions are used to represent parametrically a rational function, say $y = R(x)$, with Tchebichef behaviour, both y and x are mapped with different elliptic functions on to rectangles in the complex plane of the parameter. By using different moduli for the two elliptic functions and by having suitable multiplying constants on the parameter, the rectangle for x is arranged to be some simple integral multiple of the rectangle for y and this guarantees that a *rational* function will be generated.

The two elliptic functions required in the present application are the cd function and the dn function. The cd function is related to the sn function as the cos function is related to the sin function ; i.e., they are shifted relative to one another by a real quarter-period ($\pi/2$ or K). For real values of the variable the cd and the cos functions behave in very much the same way, oscillating between ± 1.0 and having real periods of $4K$ and 2π respectively.

The dn function is related to the sn function by $\mathrm{dn}(u) = \sqrt{(1 - k^2sn^2(u))}$; when $u = 0$, $\mathrm{dn}(u) = 1$ and when $u = K$, $\mathrm{dn}(u) = k'$: it has a real period of $2K$. There is no circular equivalent to the dn function but it degenerates into sech (u) for $k = 1$ and for real values of the variable its behaviour, between $-K$ and $+K$, is somewhat similar to that of sech (u) between $-\infty$ and $+\infty$.

The exact parametric equations are

$$\tan\frac{\phi}{2} = \sqrt{k_1}\,\mathrm{cd}\left[2n \cdot \frac{K_1}{K} \cdot u\ ;\ k_1\right] \quad \ldots \quad (11a)$$

$$\omega = \frac{1}{\sqrt{k'}} \cdot dn\left[u\ ;\ k\right] \quad \ldots \quad \ldots \quad \ldots \quad (11b)$$

With this choice of functions the prescribed band of frequencies is normalized so that the

www.americanradiohistory.com

geometric mean of the end frequencies occurs at $\omega = 1$. K and K_1 are the real quarter-periods of elliptic functions of moduli k and k_1. n is an additional constant, always integral, which specifies the degree of the rational function being generated: its physical significance will appear later.

$k' = \sqrt{1 - k^2}$ is defined as the ratio of the lower to the upper end frequency of the prescribed band. Thus, in Fig. 6

$$k' = \frac{\omega_a}{\omega_b} \quad .. \quad .. \quad .. \quad .. \quad (12)$$

k_1 is chosen relative to k so that their corresponding modular constants q_1 and q (defined as $q = \exp\left(-\pi\frac{K'}{K}\right)$ etc.) are related by

$$q_1 = q^{4n} \quad .. \quad .. \quad .. \quad .. \quad (13)$$

This relationship, together with the choice of the multiplier $2n\frac{K_1}{K}$ in Equ. (11a), arranges that the two rectangles in the u-plane, on to which the ω-plane and the $\tan\frac{\phi}{2}$-plane respectively are mapped, have the same length of side parallel to the imaginary axis but have their sides which are parallel to the real axis, in a ratio of n to 1.

The portion of the curve of $\tan\frac{\phi}{2}$ between ω_a and ω_b in Fig. 6 is generated when the parameter u is real. As u increases from 0 to K, ω decreases, by virtue of Equ. (11b) from $\frac{1}{\sqrt{k'}}$ to $\sqrt{k'}$ while at the same time the argument of the cd function in Equ. (11a) increases from 0 to $2n.K_1$ so causing $\tan\frac{\phi}{2}$ to oscillate between $\pm\sqrt{k_1}$ n times. This gives rise to the Tchebichef behaviour. The remainder of the curve is generated when u is complex and travels round the rectangular contour with vertices K, $K + j2K'$, $j2K'$, 0.

Having now defined a suitable functional form for $\tan\frac{\phi}{2}$ by means of Equs (11a) and (11b) it is next necessary to find some means of evaluating the poles and zeros of the two transfer impedances. As the poles are the negative of the zeros it will be sufficient to find the zeros, or the poles.

From the definitions of Q and P it will be seen that

$$Q + j\omega P = (A_1A_2 + \omega^2.B_1B_2) + j\omega\,(A_2B_1 - A_1B_2) = (A_1 + j\omega B_1)(A_2 - j\omega B_2) = (A_1 + pB_1)(A_2 - pB_2) \quad .. \quad .. \quad (14)$$

Now the zeros of $(A_1 + pB_1)$ are the zeros of Z_{T1} and lie in the left half of the p-plane while the zeros of $(A_2 - pB_2)$ are the poles of Z_{T2} and so lie in the right half of the p-plane. This gives a clue to the means of finding the required poles and zeros: the zeros of $(Q + j\omega P)$ must be found and then expressed in terms of the variable $p = j\omega$, whereupon the zeros of Z_{T1} are those which lie in the left half p-plane and the poles of Z_{T2} those which lie in the right half p-plane.

The zeros of $(Q + j\omega P)$ are the roots of the equation

$$Q + j\omega P = 0$$

$$\text{or} \quad \frac{\omega P}{Q} = j \quad .. \quad .. \quad .. \quad .. \quad (15)$$

and hence the roots of

$$\frac{1 - \omega P/Q}{1 + \omega P/Q} = \frac{1 - j}{1 + j}$$

$$\text{i.e.,} \quad \frac{Q - \omega P}{Q + \omega P} = -j \quad .. \quad .. \quad .. \quad (16)$$

From Equ. (11a) this is equivalent to

$$\sqrt{k_1}\,\mathrm{cd}\left[2n\frac{K_1}{K}u\,;\,k_1\right] = -j \quad .. \quad .. \quad (17)$$

The parameter values which satisfy Equ. (17) are given by

$$2n\frac{K_1}{K}u = (4\sigma + 1)K_1 + j\tfrac{1}{2}K_1' \quad .. \quad (18)$$

[σ = any integer]

By using the relationship between the various quarter-periods, namely

$$4n\frac{K'}{K} = \frac{K_1'}{K_1} \quad .. \quad .. \quad .. \quad .. \quad (19)$$

as imposed by the requirement of Equ. (13), Equ. (18) may be written as

$$u = \frac{4\sigma + 1}{2n}\cdot K + jK' \quad .. \quad .. \quad .. \quad (20)$$

Substituting this value for u into Equ. (11b) leads immediately to the required values of ω. Thus the zeros of $(Q + j\omega P)$ are given by

$$\omega = \frac{1}{\sqrt{k'}}\cdot \mathrm{dn}\left[\frac{4\sigma + 1}{2n}\cdot K + jK'\,;\,k\right] \quad .. \quad (21)$$

By a slight transformation of the elliptic function this becomes

$$\omega = -\frac{j}{\sqrt{k'}}\cdot\frac{\mathrm{cn}\left[\frac{4\sigma + 1}{2n}\cdot K\,;\,k\right]}{\mathrm{sn}\left[\frac{4\sigma + 1}{2n}\cdot K\,;\,k\right]}$$

$$\text{or} \quad p = \frac{1}{\sqrt{k'}}\cdot\frac{\mathrm{cn}\left[\frac{4\sigma + 1}{2n}\cdot K\,;\,k\right]}{\mathrm{sn}\left[\frac{4\sigma + 1}{2n}\cdot K\,;\,k\right]} \quad .. \quad (22)$$

www.americanradiohistory.com

The function $\frac{\mathrm{cn}\,[u\,;\,k]}{\mathrm{sn}\,[u\,;\,k]}$ is the elliptic analogue of the cotangent function and degenerates into it when $k = 0$. The cotangent function has a real period of π (not 2π) and similarly the function $\frac{\mathrm{cn}\,[u\,;\,k]}{\mathrm{sn}\,[u\,;\,k]}$ has a real period of $2K$, K being analogous to $\pi/2$. For real values of the variable the two functions behave in very much the same way; they are both infinite at $u = 0$ and $u = 2K$ (or π) and zero at $u = K$ (or $\pi/2$); positive from 0 to K and negative from K to $2K$. The general shape of the function is shown in Fig. 7: altering the value of k alters the ordinates whilst retaining the main characteristics.

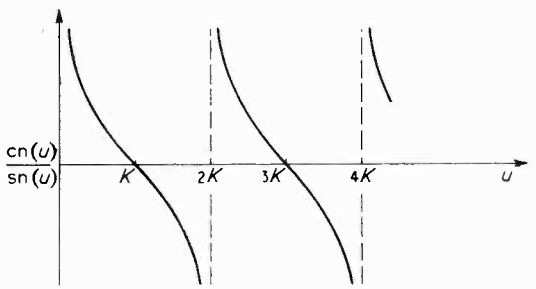


Fig. 7. *Typical curve for cn(u)/sn(u).*

If, in Equ. (22), the integral variable σ is given successively the values 0 to $(n - 1)$ inclusive, a set of n different values of p will be generated. Any other values for σ will simply repeat some member of this set, for increasing σ by n merely adds a whole period to the variable of the elliptic function.

The main result of the analysis, to be noticed at this point, is that, as the variable in the elliptic function is real for all values of σ, all the values of p which are generated will also be real. This means that the zeros of Z_{T1} and Z_{T2} are all real and that the all-pass networks could be realized physically as a tandem connection of first-degree networks.

Those values of p which are negative are immediately the zeros of Z_{T1}; reversing the sign of the positive ones gives the zeros of Z_{T2}. It will be seen, after a little examination, that the zeros of Z_{T1} interlace the zeros of Z_{T2}. Altogether, distributed between Z_{T1} and Z_{T2}, there will be n zeros and hence in the physical realization there would be n first-degree all-pass networks required. The network pair, considered as a unit, may thus be said to be 'of degree n.'

This completes the analysis necessary for the derivation of the two transfer impedances. In Sections 6 and 7 will be given methods of computing the elliptic functions of Equ. (22) and details of networks having the transfer impedances so specified. The next problem to be solved is that of finding the relationship between bandwidth, tolerance and network complexity; i.e., between k', ϕ and n.

5. Performance Relationships

Equ. (11a) shows that

$$\tan\frac{\phi}{2} = \sqrt{k_1}\,\mathrm{cd}\left[2n\frac{K_1}{K}u\,;\,k_1\right]$$

The maximum excursion of ϕ away from zero coincides with that of $\tan\frac{\phi}{2}$ away from zero. Over the band of approximation u is real and the maximum value of the cd function is ± 1. Therefore the maximum value of $\tan\frac{\phi}{2}$ is $\pm\sqrt{k_1}$. Denoting the greatest value of ϕ by ϕ_m it follows that

$$\tan\frac{\phi_m}{2} = \sqrt{k_1} \quad .. \quad .. \quad .. \quad .. \quad (23)$$

Normally $k_1 \ll 1$ and then to an extremely good approximation

$$k_1^2 \approx 16q_1 \quad .. \quad .. \quad .. \quad .. \quad (24)$$

Using this relationship together with Equ. (13) yields the required form of

$$\tan\frac{\phi_m}{2} = 2q^n \quad .. \quad .. \quad .. \quad .. \quad (25)$$

Equ. (25) relates the maximum error in the phase difference ϕ_m to the number n of elementary all-pass networks required, and the band of approximation k' which is uniquely related to q by the pair of equations

$$q = \epsilon + 2\epsilon^5 + 15\epsilon^9 + 150\epsilon^{13} + 1707\epsilon^{17} + 20910\epsilon^{21} + \quad .. \quad .. \quad .. \quad .. \quad (26a)$$

with $$\epsilon = \tfrac{1}{2}\cdot\frac{1-\sqrt{k'}}{1+\sqrt{k'}} \quad .. \quad .. \quad .. \quad (26b)$$

or $$\sqrt{k'} = \frac{1 - 2q + 2q^4 - 2q^9 + 2q^{16} - \ldots}{1 + 2q + 2q^4 + 2q^9 + 2q^{16} + \ldots} \quad (27)$$

Equs. (25) to (27) are sufficient to allow the examination of any design possibility in a matter of minutes. As a general guide, curves of ϕ_m, in degrees, against k' for several values of n have been prepared and appear in Fig. 8. The various series involved are highly convergent and in normal cases two or three terms will give adequate accuracy.

6. Computation of the Elliptic Functions

In a normal design the parameters k' and ϕ_m are prescribed from the beginning and the first operation is to find, according to Equ. (25), the smallest value of n which will meet the requirements. The curves of Fig. 8 will in most cases prove adequate for this purpose. It remains

www.americanradiohistory.com

then to find the zeros of Z_{T1} and Z_{T2} from Equ. (22) and finally two networks having these transfer impedances.

The evaluation of the elliptic functions in Equ. (22) is simplified to some extent by a consideration of the following points.

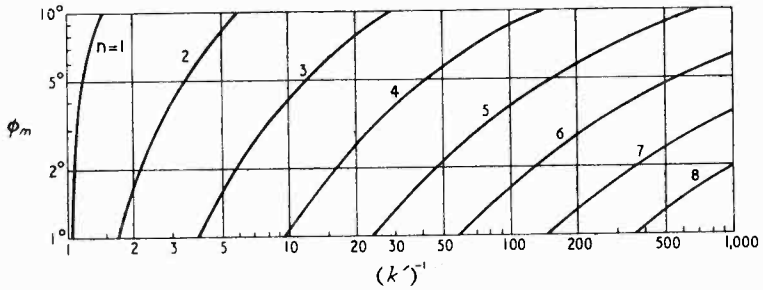


Fig. 8. *Design curves.*

First, the cn function and the sn function are related by the equation

$$\mathrm{sn}^2(u) + \mathrm{cn}^2(u) = 1 \quad \ldots \quad \ldots \quad \ldots \quad (28)$$

and so, if the appropriate sn functions are found, the corresponding cn functions can be computed from this relationship.

Secondly, once it is appreciated that the information provided by Equ. (22) simply implies that the zeros (and poles) of the two transfer impedances interlace one another, it is best to ignore the signs involved and to compute all the quantities as positive numbers; i.e., to compute the poles. They can subsequently be allocated to Z_{T1} and Z_{T2} from the knowledge that they interlace. It will be found that all the poles are included in the set of numbers defined by

$$\frac{1}{\sqrt{k'}} \cdot \frac{\mathrm{cn}\left[\frac{2\sigma+1}{2n} \cdot K\,;\, k\right]}{\mathrm{sn}\left[\frac{2\sigma+1}{2n} \cdot K\,;\, k\right]}$$

$$(\sigma = 0, 1, \ldots, n-1.)$$

where all the arguments of the sn and cn functions lie in the first quarter-period.

Thirdly, all the poles will be found to be symmetrically disposed on a logarithmic scale around unity: hence it is necessary to compute only half of them, the remainder being found by reciprocation. In addition if n is odd the middle pole will be unity and so needs no computing.

Thus it will be seen that it is necessary to find the values of

$$\mathrm{sn}\left[\frac{2\sigma+1}{2n} \cdot K\,;\, k\right]$$

for values of σ from 0 up to the largest number which satisfies the inequality $\sigma < \frac{n-1}{2}$. For instance if $n = 9$ then the σ values would be 0, 1, 2, 3.

As far as the evaluation of the functions is concerned tables do exist: however, inasmuch as the elliptic functions are virtually functions of two variables, namely the argument and the modulus, it becomes difficult even in quite bulky tables to provide other than rather coarse gradations of the variables. Hence if tables are used extensive interpolation in two variables must be expected. For this reason the writer has found it preferable to compute the functions directly via a sequence of Landen transformations from the corresponding circular functions. The computing scheme is simple, straightforward and reasonably foolproof, and not the least of its merits is that every number involved is positive. The other practical possibility for computation, that of using the theta functions, is perhaps slightly quicker but it is not so straightforward and great care must be taken with signs.

In its present application the Landen transformation is most conveniently expressed by the equations:

$$\mathrm{sn}\,[\alpha K\,;\, k] = \frac{(1+\lambda)\,\mathrm{sn}\,[\alpha \wedge\,;\, \lambda]}{1 + \lambda \cdot \mathrm{sn}^2\,[\alpha \wedge\,;\, \lambda]} \quad \ldots (29a)$$

where $\lambda = \dfrac{1-k'}{1+k'}$ $\ldots \quad \ldots \quad \ldots \quad \ldots$ (29b)

It relates two elliptic functions of different moduli, k and λ, connected by Equ. (29b), whose arguments are the same fraction of their respective quarter-periods, $\alpha.K$ and $\alpha \wedge$. λ is smaller than k. By repeating the transformation on the modulus indicated by Equ. (29b) only a few times, a point is soon reached where the modulus is vanishingly small. At this point the sn function and the sin function are near enough identical; so the procedure is to replace the sn function by

$$\sin\left[\alpha \cdot \frac{\pi}{2}\right]$$

and, by using Equ. (29a) the appropriate number of times, to retrace the path back to sn $(\alpha K\,;\, k)$.

In the present case $\alpha = \dfrac{2\sigma+1}{2n}$. The quantities $\sin\left[\dfrac{2\sigma+1}{2n} \cdot \dfrac{\pi}{2}\right]$ are independent of k and may be tabulated once and for all. Table 1 shows how the scheme is arranged. k' is known and is filled in to start with. Next λ is calculated from Equ. (29b) and λ' from the relation $\lambda^2 + \lambda'^2 = 1$. This is repeated, a new and smaller modulus being derived from λ' by Equ. (29b) and placed

www.americanradiohistory.com

below λ, and so on until a point is reached where it cannot be continued with the number of figures being used because the modulus vanishes and the complementary modulus equals unity.

TABLE I

Modulus	Complementary Modulus	1st Sn Function	2nd Sn Function
k	k'	↑	↑
λ	λ'		
↓			
Modulus Vanishes	Comp. Mod. ≈ 1.0	Sin Function	Sin Function

Then, alongside this negligible modulus, is placed the sin function of the same fraction of $\pi/2$, as is the required sn function of its quarter-period. By using Equ. (29a) the sn functions of the different moduli in the chain may be computed one after another until finally that for modulus k is reached.

Having found all the sn functions in this way the cn functions are found by Equ. (28) and then, finally, the values of the poles. This yields those poles greater than unity: the remainder, less than unity, are found by reciprocating those computed. To these is added one at unity if n is odd. The poles can now be allocated to Z_{T1} and Z_{T2} by the fact that they interlace.

It is hoped that the numerical example in Section 8 will clarify these points completely.

7. Physical Realization of Transfer Impedances

As the poles and zeros of the two transfer impedances are all real it follows, as mentioned in Section 4, that the impedances could be produced by a tandem connection of first-degree networks. In such an event each network is responsible for producing one pole-zero pair in the transfer impedance and the values of the elements, L and C, in any one network are given by

$$L = \frac{R_0}{p_r \cdot \omega_0} \qquad C = \frac{1}{p_r \cdot R_0 \cdot \omega_0} \quad \ldots \quad (30)$$

where R_0 is the characteristic impedance of the network, ω_0 the geometric mean of the end angular frequencies of the required band and p_r the normalized pole value as given by the computation of Section 6.

All the sections which are joined in tandem must have the same characteristic impedance or, if not, must form an image-matched tandem connection, preferably with the aid of matching attenuators. Also they must be correctly matched at the unparalleled ends.

Despite their simplicity first-degree networks are not the most practicable. This becomes apparent whenever an attempt is made to produce a pair of networks operating over a reasonably wide band of frequencies: then it is found that the networks which are operating over the low-frequency end of the band as 'all-pass' networks are also acting at the higher frequencies as low-pass filters with a cut-off frequency inside the prescribed band. The cause of this is the self-capacitance of the coils, and it being normally impossible to place the self-resonance of the coils in the low-frequency sections outside the complete band.

The remedy for this is to combine pairs of first-degree networks together and realize them as second-degree networks. It is then possible to place across the offending coil an intentional capacitance which can more than absorb the coil self-capacitance. In combining a pair of first-degree networks use may be made of the equivalence of Fig. 9.

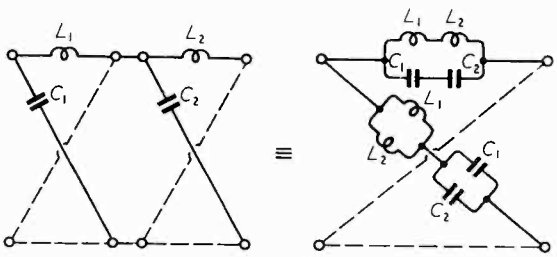


Fig. 9. *Combination of all-pass networks.*

It will be found most advantageous if the pairs which are combined are, first, the ones highest and lowest in frequency; next, the highest and lowest of what remains, and so on until either none or one is left. The impedance of the networks should then be made as low as possible, compatible with the capacitors being not unreasonably large. It may frequently turn out that different impedances are desirable for different networks in the tandem connection. In this case impedance matching attenuators should be fitted between the networks where appropriate.

Where exact equality between the magnitudes of the output voltages is required trouble may be experienced due to dissipation in the networks. This causes an attenuation component, sensibly proportional to $\frac{d\beta}{d\omega}$, in the network as a whole. As $\frac{d\beta}{d\omega}$ is not the same for each arm of the network pair the equality between the voltages tends to be destroyed. One possible remedy

www.americanradiohistory.com

is to use the equivalence indicated in Fig. 10, absorbing a suitable attenuator into each second-degree network and using one of the resistors so produced in each arm to represent the dissipation resistors. It means, that, physically, the element values in the network are changed slightly and one additional resistor is fitted to each arm.

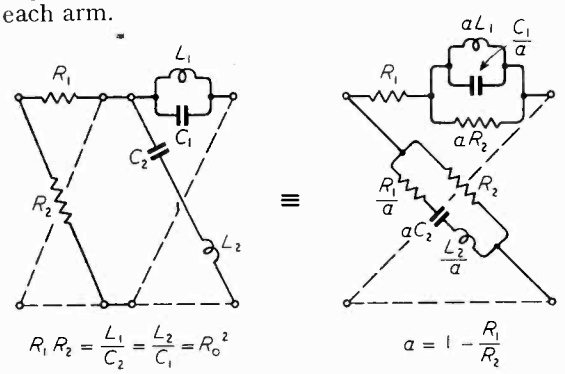


Fig. 10. *Scheme for absorbing dissipation.*

8. Numerical Example

The full significance of all the various points dealt with in the preceding sections will, perhaps, be appreciated more easily with the aid of a numerical example. The following design has been chosen with this specific end in view: it is not an example which has occurred in any particular practical problem and it should not be considered necessarily as typical. Usually practical designs are somewhat more stringent and as such are unsuitable for illustrative purposes because they become numerically rather tedious.

Consider the case of a network pair required to operate over the band of frequencies 200 c/s to 4,000 c/s and to have a phase angle tolerance of not more than $\pm 3.0°$. The first problem is to decide upon the degree of the network pair; i.e., the value of n. From Fig. 8 it can be seen that for a frequency ratio of 4,000 : 200 = 20 : 1 a network of degree 4 will provide a tolerance of $\pm 2.5°$.

This is checked by means of Equ. (25) to (27) as follows: the complementary modulus, k' is given by $k' = 200/4{,}000 = 0.05$, and hence $\sqrt{k'} = 0.2236{,}0680$. From $\sqrt{k'}$ is derived, by Equ. (26b),

$$\epsilon = \tfrac{1}{2} \cdot \frac{1 - 0.2236{,}0680}{1 + 0.2236{,}0680} = 0.3172{,}5600$$

and the sum of the first four terms of the series of Equ. (26a) gives the modular constant $q = 0.3242$. Finally from Equ. (25)

$$\tan \frac{\phi_m}{2} = 2q^{\frac{1}{4}} = 0.0221$$

and so $\phi_m = 2.5°$ thus confirming the figure obtained from the curve.

The next process is to set up the table of Landen transformations: this is shown in Table II.

First, the quantities in the k and k' columns are computed: each quantity in the k column is derived from the k' value in the row above by Equ. (29b) and each value of k' from the adjacent k by the fact that the sum of their squares equals unity. The process terminates after five operations when the value of k' is indistinguishable, to eight significant figures, from unity. Then in the same row and in the columns labelled $\mathrm{sn}[\frac{1}{8}K; k]$ and $\mathrm{sn}[\frac{3}{8}K; k]$ are placed the values of $\sin\left[\frac{1}{8} \cdot \frac{\pi}{2}\right]$ and $\sin\left[\frac{3}{8} \cdot \frac{\pi}{2}\right]$ respectively. By using Equ. (29a) and with the value of modulus appropriate to each row, the elliptic function in the row above is obtained. This is repeated until row 1 is filled in.

The Landen transformations give

$$\mathrm{sn}[\tfrac{1}{8}K; k] = 0.4990{,}8787$$
$$\mathrm{sn}[\tfrac{3}{8}K; k] = 0.9284{,}7394,$$

by Equ. (28)

$$\mathrm{cn}[\tfrac{1}{8}K; k] = 0.8665{,}5138$$
$$\mathrm{cn}[\tfrac{3}{8}K; k] = 0.3713{,}9755$$

and so

$$\frac{1}{\sqrt{k'}} \frac{\mathrm{cn}[\frac{1}{8}K; k]}{\mathrm{sn}[\frac{1}{8}K; k]} = 7.7648{,}3617$$

$$\frac{1}{\sqrt{k'}} \frac{\mathrm{cn}[\frac{3}{8}K; k]}{\mathrm{sn}[\frac{3}{8}K; k]} = 1.7888{,}9278$$

These are two of the four poles: the other two are the reciprocals of these. Thus the four poles are

1. 7.7648,3617; 3. 0.5590,0499;
2. 1.7888,9278; 4. 0.1287,8571.

Poles No. 1 and No. 3 belong to network No. 1, while poles No. 2 and No. 4 belong to network No. 2. It is now a straightforward matter to compute from Equ. (30) the component values in the elementary first-degree networks and to combine them with the aid of Fig. 9. A suitable value for R_0 is 1,200Ω; and ω_0 is given by $2\pi\sqrt{200 \times 4{,}000} = 5619.85$.

TABLE II

	k	k'	$sn[\frac{1}{8}K; k]$	$sn[\frac{3}{8}K; k]$
1		0.0500,0000	0.4990,8787	0.9284,7394
2	0.9047,6191	0.4259,1770	0.2807,0024	0.7093,8393
3	0.4026,0549	0.9153,7360	0.2034,6320	0.5725,0008
4	0.0441,8271	0.9990,2347	0.1951,8200	0.5557,5785
5	0.0004,8850	0.9999,9988	0.1950,9033	0.5555,7025
6	0.0000,0006	$1.0 - 18 \times 10^{-16}$	0.1950,9032	0.5555,7023

www.americanradiohistory.com

The final circuit is shown in Fig. 11.

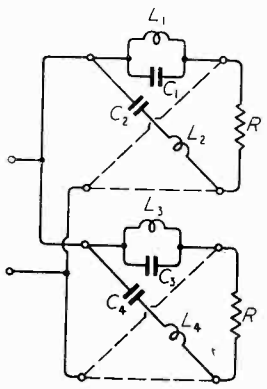


Fig. 11. *Circuit for numerical example.*

The component values are

$L_1 = 0.4094$ H
$L_2 = 0.0236$ H
$L_3 = 1.7774$ H
$L_4 = 0.1113$ H
$C_1 = 0.0178\ \mu$F
$C_2 = 0.2844\ \mu$F
$C_3 = 0.0773\ \mu$F
$C_4 = 1.2343\ \mu$F
$R = 1200\Omega$

Acknowledgments

This paper is published by permission of the Engineer-in-Chief of the P.O. Engineering Department, in the research laboratories of which the investigation was carried out. The writer is deeply indebted to Dr. J. M. Linke both for his many suggestions in connection with the work and also for his constructive criticism of the manuscript.

REFERENCES

[1] U.K. Patent No. 547,601.

[2] Dome, R. B.: "Wideband Phase Shift Networks," *Electronics*, Dec. 1946, p. 112.

[3] Luck, D. G. C.: "Properties of some Wideband Phase Splitting Networks," *Proc. Inst. Radio Engr*, Feb. 1949, p. 147.

[4] Bode, H. W.: "Network Analysis and Feedback Amplifier Design," D. Van Nostrand, New York, 1945.

RC-COUPLED POWER STAGE

Conditions for Maximum Output

By M. G. Scroggie, B.Sc., M.I.E.E.

FOR drawing the a.c. power output from a valve it is usual to employ a choke or transformer coupling whose d.c. resistance is small and a.c. impedance large compared with either valve or load impedance. By so doing, the device for excluding d.c. from the load can, to a first approximation, be neglected when calculating the power relationships of the stage.

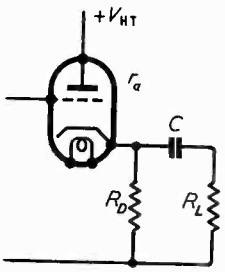


Fig. 1. *The type of coupling considered is shown here as applied to a cathode follower, but the analysis also covers connection in the anode circuit. C is assumed to have negligible impedance at the working frequency, so the effective a.c. load, denoted by R_A, is R_D and R_L in parallel.*

There are various familiar expressions, based on this assumption, giving the load resistance for maximum output power as a function of the valve characteristics, depending on the particular conditions laid down; e.g., constant grid drive, constant distortion, etc.

When the valve feed or coupling device is, like the load, a resistance (as for example in Fig. 1) both enter into the problem; and their values for maximum power in the load do not appear to be well known. The conditions for which they will now be derived are those giving maximum 'undistorted' power output of a sine wave; viz., the grid bias and grid drive are so adjusted as to swing the valve exactly between the points at which grid current starts and anode current is cut off. The valve characteristics are assumed linear. The output circuit may be connected on either anode or cathode side of the valve, but in the latter case the ideal conditions which have just been assumed are approximated more closely by reality. (With no negative feedback, distortion usually becomes appreciable before anode-current cut-off is reached). It will be shown that under these conditions the maximum power, P_{max}, is obtained in R_L when

$$R_L = r_a$$
$$\text{and } R_D = \sqrt{2}r_a$$

and that when these conditions are fulfilled

$$P_{max} = V^2/93r_a$$

where r_a is given by the nearest linear approximation to the I_a/V_a curve at the threshold of grid current (BE in Fig. 2) and V is the anode supply voltage less the voltage at the point where the r_a line cuts the V_a axis.

Fig. 2 also shows the R_D line, drawn between the point C marking the anode supply voltage, V_{HT}, and the r_a line at D. To fulfil the specified conditions, the working point A, must be so placed on the R_D line as to bisect the a.c. load

MS accepted by the Editor, July 1949

www.americanradiohistory.com

line, EF. Since R_A, the a.c. load, is R_D and R_L in parallel,

$$R_L = \frac{R_D R_A}{R_D - R_A} \quad .. \quad .. \quad .. \quad .. \quad (1)$$

Denoting the peak voltage across the load by v, and the power output for any values of R_D and R_L by P,

$$P = v^2/2R_L \quad .. \quad .. \quad .. \quad .. \quad (2)$$

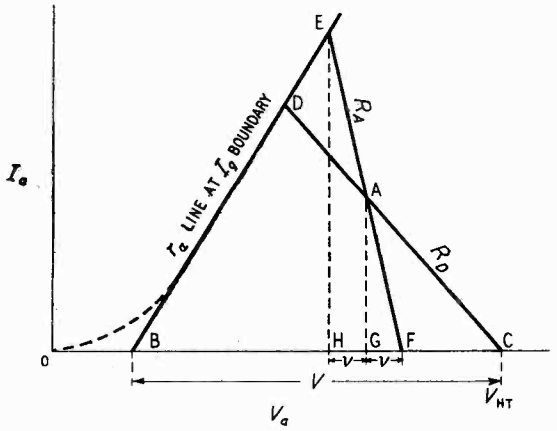


Fig. 2 (above). *A.c. and d.c. load lines on the linearized valve characteristic plane.*

Fig. 3. (right). *Calculated power output (P) in R_L as a function of R_L and R_D, generalized as regards r_a and V.*

From the geometry of Fig. 2 and the conditions specified

$$v = \frac{HF}{2} = \frac{BH}{2} \cdot \frac{R_A}{r_a}$$

$$\text{So } BH = 2v \cdot \frac{r_a}{R_A} \quad .. \quad .. \quad (3)$$

$$\text{Also } v = GC \cdot \frac{R_A}{R_D}$$

$$= [V - (BH + v)] \frac{R_A}{R_D}$$

Substituting (3)

$$v = [V - (2v \frac{r_a}{R_A} + v)] \frac{R_A}{R_D}$$

$$= \frac{VR_A}{R_A + R_D + 2r_a}$$

$$= \frac{VR_D R_L}{R_D R_L + (R_D + 2r_a)(R_D + R_L)} \quad (4)$$

and, from (2)

$$P = \left[\frac{VR_D}{R_D R_L + (R_D + 2r_a)(R_D + R_L)}\right]^2 \frac{R_L}{2} \quad (5)$$

Differentiating P with respect to R_L and equating to zero,

$$R_L = \frac{R_D (R_D + 2r_a)}{2(R_D + r_a)} \quad .. \quad .. \quad (6)$$

which gives a maximum for P.

Substituting (6) in (5),

$$P = \frac{V^2 R_D}{16(R_D^2 + 3r_a R_D + 2r_a^2)} \quad .. \quad .. \quad (7)$$

Differentiating P with respect to R_D and equating to zero,

$$R_D = \sqrt{2} r_a \quad .. \quad .. \quad .. \quad .. \quad (8)$$

which is the value giving maximum P.

Substituting in (6),

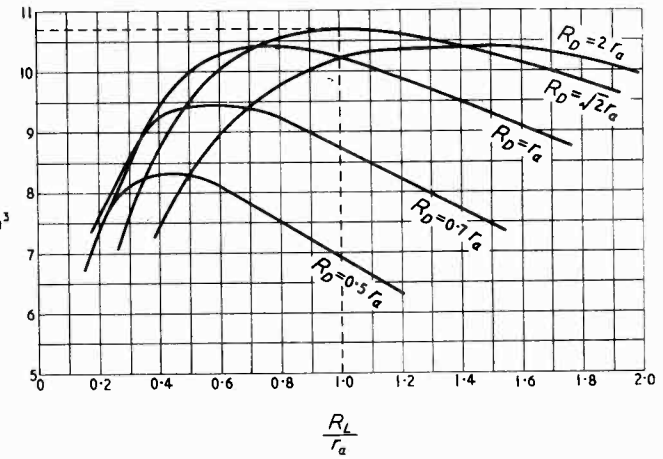
$$R_L = r_a \quad .. \quad .. \quad .. \quad .. \quad .. \quad (9)$$

Substituting (8) in (7) to find the maximum power,

$$P_{max} = \frac{V^2}{(48 + 32\sqrt{2}) r_a} = \frac{V^2}{93.26 r_a} \quad .. \quad .. \quad (10)$$

$$\text{Also } R_A = \frac{r_a}{1 + 1/\sqrt{2}} = 0.586 r_a \quad .. \quad .. \quad (11)$$

Using equation (5), the relationship between output power and load resistance is shown in Fig. 3 for several values of R_D, from which it can be seen that the optimum values of R_L and R_D



are not at all critical. If both are multiplied or divided by any factor not exceeding $\sqrt{2}$, the consequent loss of power does not exceed 3 per cent.

The measured power, using a triode-connected SP61 valve as a cathode follower, with $r_a = 6.6$ kΩ and $V = 210$V, agreed very satisfactorily with the foregoing theory, the calculated P_{max} (72 mW) being obtained close to the point at which distortion of the sinusoidal waveform began to be clearly visible on an oscilloscope.

www.americanradiohistory.com

TRANSIENT RESPONSE OF A REGULATOR CHAIN

By H. Jefferson, M.A., A.M.I.E.E.

MANY communication systems depend for their satisfactory operation upon the functioning of automatic level regulators, which may be controlled either by the actual signal or by a special pilot signal. One example of such an automatic regulator is the automatic gain-control circuit of a radio receiver, in which the carrier may be regarded as a pilot signal, in contrast to the use of the speech modulation in the constant-volume amplifier which follows the receiver in a single-sideband radio-telephone circuit. Carrier telephone systems on lines and cables, and microwave radio links, may include a number of repeater stations, each incorporating a regulator to maintain the signal at the correct level before transmitting it to the next section.

Level regulators are not instantaneous in action but have a finite operating time. If a change of level occurs at the input to a regulator there is, in consequence, a transient change of level at the output. The effect of this transient on succeeding regulators in the chain is of considerable importance, for it is known that a transient of damped oscillatory character can be set up at points along the chain. These oscillations of level are, of course, associated with oscillations of gain in the amplifiers of which the regulators form a part. Not only are such changes of gain disturbing to the users of the circuit, but should the amplitude of the gain oscillations be excessive a four-wire circuit may oscillate due to excessive loop gain. Should the oscillations of level be excessive, distortion due to overloading may result. The frequency of the oscillations is usually very low, periods of the order of 1 to 10 seconds being common. The frequency is dependent on the characteristics of the level-regulating system.

In the discussion which follows certain assumptions are made regarding the control law of the regulators. These assumptions are chosen to provide a simple mathematical solution, but the characteristics chosen do not deviate too widely from those of practical regulators.

A regulator circuit consists essentially of two components: a variable attenuator in the signal path, and a control or operator circuit bridged across the signal path. The operator circuit measures the level of the signal and produces a voltage or current which is fed to the variable attenuator to adjust the loss. When the regulator is required to maintain a constant output level it is preferable to bridge the operator circuit across the output, thus ensuring a monotonic input-output characteristic. Expander circuits, which belong to the same general class of level-control devices, are best designed with the operator bridged across the input. The two basic units described above are normally supplemented by an amplifier, which provides the required output level.

The portion of the communication system between the output of one level indicator and the output of the next in the chain, which will include a length of cable or a single radio path and one level regulator, will be referred to as a repeater section, and it will be assumed that the regulator is adjusted to provide unit amplification in the normal condition, measured over a repeater section. This is often referred to as zero equivalent.

All gains and losses will be referred to in logarithmic units; which unit is used is unimportant until numerical calculation is involved.

It is assumed that the operator circuit measures the average value of output level, and uses this to control the attenuation. Normally the output level is zero; that is, equal to the reference level for the logarithmic units. (For example, the output level may be 0 db for 1 milliwatt). In this event there is no output from the operator circuit. Averaging may be effected either by rectifying the output voltage and storing charge in a capacitor, or by applying the controlling signal to a thermal element, the temperature of which affects the attenuation.

With these assumptions the characteristics of a repeater section become

$$N_2 = A + N_1 \quad \ldots \quad \ldots \quad \ldots \quad (1)$$

$$A = A_0 - k\int_0^t N_2 dt \quad \ldots \quad \ldots \quad (2)$$

where

N_1 = input level in logarithmic units
N_2 = output level
A_0 = 'standard' gain
A = actual gain of the section at time t
k = a constant of the system

MS accepted by the Editor, April 1949

www.americanradiohistory.com

Writing

$\int_0^t dt = 1/p$, equations (1) and (2) lead to

$$N_2 = A_0 - \frac{k}{p} N_2 + N_1$$

or

$$\frac{p+k}{p} N_2 = A_0 + N_1 \quad .. \quad .. \quad .. \quad (3)$$

The standard gain A_0 is to be zero, so that we have

$$N_2 = \frac{p}{p+k} N_1 \quad .. \quad .. \quad .. \quad (4)$$

If the input level is suddenly increased by one unit (say 1 db) the output level is given by

$$N_2 = \frac{p}{p+k} \mathbf{1} \quad .. \quad .. \quad .. \quad (5)$$

the solution of which is known to be

$$N_2 = e^{-kt} \qquad t>0 \quad .. \quad .. \quad .. \quad (6)$$

For a single repeater section, therefore, a sudden increase in input level produces instantaneously an equal increase in output level. The output level then falls exponentially towards the standard value with time constant $1/k$.

For two repeater sections in tandem, we have, for the first,

$$N_2 = \frac{p}{p+k} N_1$$

and for the second section

$$N_3 = \frac{p}{p+k} N_2$$

where N_3 is the output from the second section, and N_2 the output from the first section is also the input to the second. Clearly, then,

$$N_3 = \left(\frac{p}{p+k}\right)^2 N_1 \quad .. \quad .. \quad .. \quad (7)$$

Considering a step input as before, with $N_1 = \mathbf{1}$

$$N_3 = \left(\frac{p}{p+k}\right)^2 \mathbf{1} \quad .. \quad .. \quad .. \quad (8)$$

$$= p \cdot \frac{p}{(p+k)^2} \mathbf{1}$$

$$= p(te^{-kt})$$

$$= (1 - kt)e^{-kt} \quad .. \quad .. \quad .. \quad (9)$$

A response of this sort has a zero at $t = 1/k$ and a turning point at $t = 2/k$, when the value is $-e^{-2} = -0.135$. It indicates that two repeater sections in tandem, when the input level is raised suddenly, give an output which at first rises by an amount equal to the rise in input. It then falls, passing through the reference level after a time $1/k$ and reaches a minimum at a time $2/t$, after which the level rises monotonically to the reference level. The amplitude of the overswing is 13.5 % of the input step amplitude.

Extending the equations to apply to n identical repeater sections, we have :

$$\left.\begin{aligned} &\text{for the first} \quad N_2 = \frac{p}{p+k} N_1 \quad .. \\ &\text{for the second} \quad N_3 = \frac{p}{p+k} N_2 \quad .. \\ &\cdots\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots\cdots \\ &\text{for the } n^{\text{th}} \quad N_{n+1} = \frac{p}{p+k} N_n \quad .. \end{aligned}\right\} \quad (10)$$

so that

$$N_{n+1} = \left(\frac{p}{p+k}\right)^n N_1 \quad .. \quad .. \quad (11)$$

Assuming, as before, that N_1 is a unit step input, and rearranging (11) slightly :

$$N_{n+1} = p^{n-1} \frac{p}{(p+k)^n} (1) \quad .. \quad .. \quad (12)$$

The solution of this is known, being

$$N_{n+1} = \frac{d^{n-1}}{dt^{n-1}} \cdot \frac{t^{n-1}}{(n-1)!} e^{-kt} \quad .. \quad (13)$$

For a small number of repeater sections this can be evaluated by carrying out the indicated differentiation. A more formal solution, however is obtained by using the Laguerre polynomial.[1]

Then

$$N_{n+1} = \frac{e^{-kt}}{(n-1)!} L_{n-1}(kt) \quad .. \quad .. \quad (14)$$

The properties of the Laguerre polynomial $L_n(z)$ are known, and tabulated values have been published, although they are not easily accessible. The function has all its zeros, a finite number, between zero and infinity, and they are all real. In consequence, the output level from a regulator chain to which a change of input level is made oscillates a finite number of times about zero level and then decays monotonically to zero level. A system having n repeater sections will have an output response characteristic which crosses the reference level axis $(n-1)$ times. The time scale depends only on k, which is a characteristic of the regulators, but the amplitude is independent of k and depends only on the magnitude of the initial disturbance. The actual form can be seen from the figure, which shows the results for up to five repeater sections. It will be noted that the oscillations become more rapid and larger in amplitude as the number of repeater sections is increased.

The condition that k shall be the same for all regulators in the chain is one normally set by the equipment designer. If this limitation is

www.americanradiohistory.com

not made, equations (10) become :

$$N_2 = \frac{p}{p + k_1} N_1 ; N_3 = \frac{p}{p + k_2} N_2 ;$$

$$N_{n+1} = \frac{p}{p + k_n} N_n \quad .. \quad .. \quad (15)$$

so that

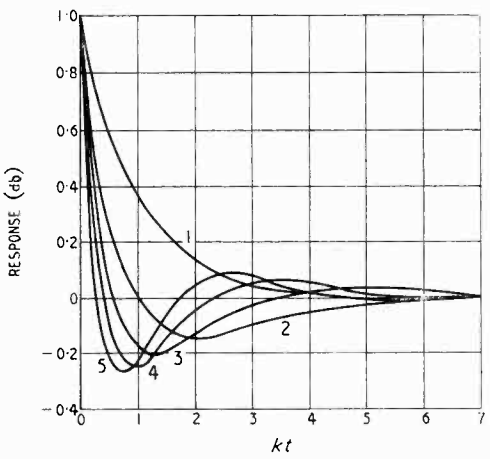
$$N_{n+1} = \frac{p^n}{\prod_{\nu=r}^{n} (p + k_\nu)} N_1 \quad .. \quad .. \quad (16)$$

This expression can be expanded into partial fractions, which then give the solution, for a unit step input :

$$N_{n+1} = \sum_{\mu} \frac{(-k\mu)^{n-1}}{\prod_{\mu \neq \nu} (k_\nu - k_\mu)} \exp(-k_\mu t) \quad .. \quad (17)$$

for $t > 0$.

Expressions of this type are encountered in studies of sequences of radioactive disintegrations[2]. As the regulators for which this study was carried out cannot conveniently be constructed to operate with different values of k, no further investigation of (17) has been made.



Relative output level from a system of (n − 1) regulated repeaters following a 1-db increase of input level.

In a practical regulator the integration process is imperfect owing to the presence of leakage. A thermally-controlled system will lose heat by radiation and convection whenever its temperature deviates from the rest condition temperature : a capacitance will lose charge by leakage. If the discharge time constant is $1/m$, the characteristic equation for a single repeater section becomes :

$$N_2 = \frac{p + k}{p + m + k} N_1 \quad .. \quad .. \quad .. \quad (18)$$

For n identical repeater sections, equation (11) becomes :

$$N_{n+1} = \left(\frac{p + m}{p + m + k}\right)^n N_1 \quad .. \quad .. \quad (19)$$

For a unit step change of input level, the output is

$$N_{n+1} = \left(\frac{p + m}{p + m + k}\right)^n \mathbf{1} = \left[p^{n-1} + m\,(n-1)\,p^{n-2} + \frac{m^2(n-1)(n-2)}{2!} p^{n-3} + \ldots\ldots + m^{n-1}\right] \frac{p}{(p+k+m)^n} \cdot \mathbf{1} \quad (20)$$

$$= [p^{n-1} + m\,(n-1)\,p^{n-2} \ldots + m^{n-1}] \frac{t^{n-1}}{(n-1)!} \cdot e^{-(k+m)t} \quad .. \quad .. \quad (21)$$

$$= \left[\frac{L_{n-1}\{(k+m)t\}}{(n-1)!} + mt \frac{L_{n-2}\{(k+m)t\}}{(n-2)!} + \ldots + m^{n-1} \cdot t^{n-1}\right] e^{-(k+m)t} \quad .. \quad (22)$$

The three solutions derived in this paper provide a way of estimating the effect of the amplitude oscillations which may occur in a practical system incorporating a number of automatic level regulators. They are also of value in showing how the amplitude-oscillation frequency depends on the number of regulators, an effect which is of importance if a final high speed regulator is to be used to eliminate these oscillations. It can be seen from the figure, for example, that after five repeater sections the apparent time constant has been reduced to about one-fifth of that of a single regulator. Deviations from the simple assumptions made do not normally produce any qualitative change in the behaviour of the system.

Acknowledgment

The author's thanks are due to Telefonaktiebolaget L. M. Ericsson for permission to publish this note, and to M. Bohman and Dr. Vos for advice and criticism.

REFERENCES

[1] "Methods of Mathematical Physics," H. and B. S. Jeffreys, Cambridge, 1946. Page 586.
[2] *loc. cit.*, p. 232.

www.americanradiohistory.com

INDUCED GRID NOISE

By R. L. Bell, B.Sc., Ph.D.

SUMMARY.—It is shown that under favourable conditions there is an exact relationship between mean-square noise currents induced at the grid of a triode valve and the 'space-charge' component of input capacitance measurable at that point. In practice this provides an approximate relation useful for estimating the total mean-square induced noise from other measurements.

Techniques of grid-noise measurement are discussed and attention drawn to the occurrence of various inherent errors. Results are presented of measurements of induced grid noise in microwave triodes at 200 Mc/s.

It is concluded that available three-halves power-law treatments are inadequate to explain the induction of shot noise at a control grid, but that values of space-charge admittance measured on the valve together with other low-frequency measurements may be made to yield much closer estimates of induced grid noise in any structure, for moderate transit angles.

1. Introduction

IT is widely accepted that the shot-effect fluctuations 'induced' at the control grid are of importance in determining the signal/noise performance of amplifying valves at the higher frequencies. The precise effect of this grid noise, however, is difficult to estimate since reliable methods of calculation are not yet available and measurement of the effects on commercial valves is a prohibitively laborious process. There is evidently required some method of estimating the induced grid noise in terms of other and more readily measurable phenomena.

It is shown in the following that, for small transit angles, the noise induced at the grid is in fact intimately related to the component C_t of input capacitance at that grid traceable to the presence of space charge in the valve. This C_t is *measurable* as the increase in input capacitance of the valve between the cut-off and specified working conditions, provided that both anode and cathode are effectively earthed at the frequency at which the measurement is made.

A random noise current $i(t)$ is conveniently referred to in terms of its 'frequency spectrum' $g(\omega)$ and its 'power spectrum' $s(\omega)$, where $\omega = 2\pi \times$ frequency will be referred to hereafter as 'frequency.' Suppose we are given a definite $i(t)$ in the time interval $(0, T)$, then for that interval

$$g(\omega) = \int_0^T i(t)e^{-j\omega t}\,dt \qquad \ldots \qquad (1)$$

is defined as its frequency spectrum, and

$$s(\omega) = \lim_{T\to\infty} \frac{2|g(\omega)|^2}{T} \qquad \ldots \qquad (2$$

its power spectrum. The factor 2 enters by virtue of the fact that $|g(\omega)|^2$ is an even function of ω and thus that values of $\omega > 0$ only need be considered.

If we suppose $i(t)$ to flow through unit resistance the time-mean power dissipated will be $\overline{i^2(t)}$. The quantity

$$d\overline{i^2} = s(\omega)\,d\omega$$

involving the power spectrum $s(\omega)$ may be regarded as the contribution to the total power $\overline{i^2(t)}$ attributable to the Fourier components of $i(t)$ having frequencies between ω and $\omega + d\omega$.

If for a triode, $s_a(\omega)$ and $s_g(\omega)$ are the power spectra of anode and induced grid noise respectively, $Y_1(\omega)$ $[\approx j\omega C_t]$ the component of grid input admittance due to space charge, and $g_m(\omega)$ the high-frequency mutual conductance, it is shown in Section 3 below that

$$s_g(\omega) = G^2 \left|\frac{Y_1(\omega)}{g_m(\omega)}\right|^2 s_a(\omega) \qquad \ldots \qquad \ldots \qquad (3)$$

The value of the factor G^2 depends on geometrical and emission conditions inside the valve and is undetermined, except that it is known to lie in the vicinity of unity.

In general, the case $G^2 = 1$ corresponds to the 'ideal valve' in which all electrons are emitted from the cathode with identical normal velocities, and in which the grid is approximated to by a continuous but permeable electrode, field variations parallel to the valve electrodes being eliminated. Since it is not possible to calculate G^2 upon the removal of either of these restrictions, it must in effect be assumed that both conditions hold. Attempts at estimating G^2 from subsidiary measurements on the valves investigated have not proved very successful, and in practice calculations are made using the approximate value $G^2 = 1$.

2. Grid Current and Admittance

We are interested in the short-circuit grid current in a triode in which anode and cathode are connected to a common earth point by leads of negligible impedance. Results may then be derived in terms of the injected currents associated with the 'nodal' method of circuit analysis[1], and the general case of impedances in the leads may be had therefrom by the application of circuit theory. Thus an assumption of zero lead impedances implies no loss of generality.

The relation (2) between the frequency and power spectra of $i(t)$ may be used to deduce the

MS accepted by the Editor, May 1949.

www.americanradiohistory.com

following result. Given that a random fluctuation current $i_{12}(t)$ of power spectrum $s_a(\omega)$ flowing in one branch of a (linear) network gives rise to, and is thereby completely correlated with, a fluctuation current $i_{34}(t)$ of spectrum $s_b(\omega)$ in some other branch, the relation between s_a and s_b at a given frequency ω may be obtained by considering the behaviour of two sinusoidal currents $I_{12}(\omega)$ and $I_{34}(\omega)$, analogous to i_{12} and i_{34} and flowing in the specified branches. One has in fact that

$$\frac{s_b(\omega)}{s_a(\omega)} = \left|\frac{I_{34}(\omega)}{I_{12}(\omega)}\right|^2 \quad .. \quad .. \quad .. \quad (4)$$

as might be expected.

If we suppose that a sinusoidal current $I_{12}(\omega)$ flows in the branch joining nodes 1 and 2 of a valve circuit and gives rise, by any mechanism whatever, to the current I_{34} in another branch, the fundamental assumption of all small-signal valve analyses may be made; viz., that the two currents are linearly related:

$$I_{34}(\omega) = \rho(\omega)\, I_{12}(\omega) \quad .. \quad .. \quad .. \quad (5)$$

where $\rho(\omega)$ is some complex circuit parameter. Suppose now an impedanceless signal generator of voltage $V_{34}(\omega)$ to be applied between terminals 3 and 4. This will give rise to a current

$$I_{12} = Y_T(\omega).\ V_{34}$$

where $Y_T(\omega)$ is the transfer admittance relating I_{12} to V_{34}. The resultant current flowing between 3 and 4 will therefore be expressible in the form

$$I_{34} = Y_{oi}\, V_{34} + \rho\, Y_T\, V_{34}$$

the input admittance between 3 and 4 being

$$Y_i = Y_{oi} + \rho Y_T$$

Whence in terms of admittances

$$\rho = \frac{Y_i - Y_{oi}}{Y_T} \quad .. \quad .. \quad .. \quad .. \quad (6)$$

Thus we have for fluctuation power spectra

$$s_b(\omega) = \left|\frac{Y_i - Y_{oi}}{Y_T}\right|^2 s_a(\omega) \quad .. \quad .. \quad (7)$$

where $s_a(\omega)$ is the spectrum of some fluctuation internal to the network, s_b the spectrum of the observable fluctuation due solely to the source s_a, Y_i the input admittance at the point of observation and Y_T a transfer admittance from that point to the above internal branch of the network.

Since ρ is finite [$\rho < \infty$], from (6) Y_{oi} is the value of Y_1 observable when $Y_T = 0$. In general Y_T will be some function $Y_T(\rho)$ of the 'feedback coefficient' ρ and will be expressible in the form

$$Y_T(\rho) = Y_T(0) + \rho Y_T'(0) + \frac{\rho^2}{2} Y''_T(0) \ldots\ldots$$

Substituting this in (7),

$$\frac{s_b(\omega)}{s_a(\omega)} = \left|\frac{Y_i - Y_{oi}}{Y_T(0)}\right|^2 \left|1 + \rho\frac{Y_T'(0)}{Y_T(0)} + \frac{\rho^2}{2}\frac{Y_T''(0)}{Y_T(0)} \ldots\ldots\right|^{-2} \quad (8)$$

The values of the terms $\rho\,\frac{Y_T''(0)}{Y_T(0)}$, $\frac{\rho^2}{2}\frac{Y_T''(0)}{Y_T(0)}$, etc, encountered in the problems to be discussed are such that their contribution to the second bracket is in the nature of a small 'correction' only. Neglecting this effect, we have to a close approximation in the presence of feedback

$$s_b(\omega) = \left|\frac{Y_i - Y_{oi}}{Y_T(0)}\right|^2 s_a(\omega) \quad .. \quad .. \quad (9)$$

Here $Y_T(0)$ is the value of Y_T calculable in the absence of feedback ($\rho = 0$).

In calculating the fluctuation power spectrum appearing at a valve grid, we let s_b and s_a be the power spectra of the short-circuit current fluctuations in the grid-earth and cathode-anode branches respectively. The value $Y_T(0)$ of the transfer admittance Y_T in the absence of feedback to the grid is simply the high-frequency mutual conductance $g_m(\omega)$ of the valve. Y_i is the input admittance at the valve grid under actual operating conditions and Y_{oi} the value of Y_i when $g_m = 0$; i.e., at cut off. The difference $Y_i - Y_{oi}$ may be expressed as an 'active' component of grid input admittance Y_1.

Hence, as a close approximation for the grid fluctuation spectrum $s_g(\omega)$, we have

$$s_g(\omega) = \left|\frac{Y_1(\omega)}{g_m(\omega)}\right|^2 s_a(\omega) \quad .. \quad .. \quad (10)$$

where the admittance Y_1 itself arises from the effects of feedback to the grid. Examples of such admittances are the 'space charge' component of input capacitance at a grid due to finite electron-transit time, and the input conductance at a grid due to anode and cathode lead-inductance effects.

3. The Distributed Case

In this Section we investigate the effect on the expression (10) of a variation of the feedback coefficient ρ to the grid, with say emission energy and trajectory of the electron. The discussion will be relevant only to the case of the 'induced' component of grid noise (i.e., that due to electron transit-time effects) and will clearly *not* apply to external 'circuit' effects such as the lead-inductance feedback grid noise, since all electron groups are treated indistinguishably by external feedback mechanisms.

It is easily shown that for a sufficiently negative grid the potential minimum in a triode will attach itself to the grid wires and approach the cathode only in the regions between, as shown in Fig. 1.

Considering the electrons of emission energy λ which finally arrive at the anode from a point z on the cathode of Fig. 1; one of these electrons will cause an anode current which is a func-

www.americanradiohistory.com

tion of λ, z, and time t, $i_a(t, \lambda, z)$ and will at the same time induce a grid-current pulse $i_g(t, \lambda, z)$. These two currents, being due to the single event of the transit of a particular electron, will be completely correlated. Taking the Fourier transform (deriving frequency spectra) of these currents gives the two harmonic 'currents' $I_a(\omega, \lambda, z)$ and $I_g(\omega, \lambda, z)$ which are completely defined, one implicitly in terms of the other. The conditions for the application of the admittance relation (10) are thus satisfied by the transit of every electron of this type, and some form of (10) will therefore apply to the grid noise induced by the 'transmitted' type of electron.

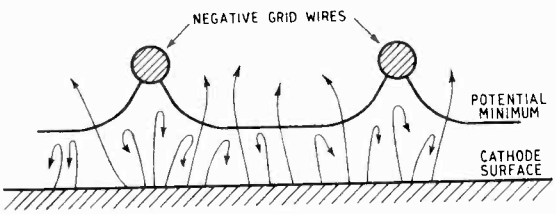


Fig. 1. *Shape of potential minimum and electron trajectories for a negative grid.*

Considering next an electron which fails to cross the potential minimum and returns to the cathode in the region midway between grid wires, it will momentarily effect the space-charge distribution in the triode, giving rise to an anode current $i_a(t, \lambda, z)$. The net effect will induce a grid current $i_g(t, \lambda, z)$ and the two currents will specify each other completely. We are led to the result that the admittance relation applies equally to electrons of this group. Moreover, it is clear that there is no essential difference between the type of reflected electron just discussed and that which is emitted from that part of the cathode immediately under a grid wire, for all of these will have some effect on the anode current.

We have then that for all λ and z there is an induced grid current $i_g(t, \lambda, z)$ with a spectrum $I_g(\omega, \lambda, z)$, anode currents $i_a(t, \lambda, z)$ and $I_a(\omega, \lambda, z)$, and that the corresponding anode and grid currents are completely correlated. We may, therefore, attempt an application of (10) to the total induced grid noise.

If $\mu(\lambda, z)\, d\lambda\, dz$ be the rate of emission of electrons of normal component of emission energy in the range $(\lambda, \lambda + d\lambda)$ from a region $(z, z + dz)$ of the cathode, and the frequency spectrum of the grid-current pulse due to one of these electrons be $I_g(\omega, \lambda, z)$, then following the method of a previous paper[2] the power spectrum of the total induced grid noise is given by a repeated application of Campbell's theorem:

$$s_g(\omega) = 2\int_\lambda d\lambda \int_z \mu(\lambda, z) \left| I_g(\omega, \lambda, z) \right|^2 dz \qquad (11)$$

Clearly, if $Y_1(\omega, \lambda, z)\, d\lambda\, dz$ and $g_m(\omega, \lambda, z)\, d\lambda\, dz$ are respectively the contributions to grid admittance and mutual conductance made by electrons in the above group $d\lambda\, dz$, we can write

$$I_g(\omega, \lambda, z) = \frac{Y_1(\omega, \lambda, z)\, d\lambda\, dz}{g_m(\omega, \lambda, z)\, d\lambda\, dz} I_a(\omega, \lambda, z) \qquad (12)$$

where $I_a(\omega, \lambda, z)$ is the frequency spectrum of the anode-current pulse due to an electron of that group. It is to be noted that the angle of emission of the electron (or alternatively transverse emission energy) can be taken into account by a slight extension of the procedure here, and will therefore be ignored.

One has then from (11) and (12) that

$$s_g(\omega) = 2\int_\lambda d\lambda \int_z \mu(\lambda z) \left| \frac{Y_1(\omega,\lambda,z)}{g_m(\omega,\lambda,z,)} \right|^2 |I_a(\omega,\lambda,z)|^2 dz \qquad (13)$$

and similarly for the anode noise

$$s_a(\omega) = 2\int_\lambda d\lambda \int_z \mu(\lambda, z) \left| I_a(\omega,\lambda,z) \right|^2 dz \qquad (14)$$

For the 'ideal valve' with single-valued fields and velocities, $Y_1(\omega, \lambda, z)$ and $g_m(\omega, \lambda, z)$ no longer vary with z, and only a very narrow band of values of λ is possible, hence

$$\left| \frac{Y_1(\omega,\lambda,z)}{g_m(\omega,\lambda,z)} \right|^2 = \frac{\left| \int_z \int_\lambda Y_1(\omega,\lambda,z)\, d\lambda\, dz \right|^2}{\left| \int_z \int_\lambda g_m(\omega,\lambda,z)\, d\lambda\, dz \right|^2} = \left| \frac{Y_1(\omega)}{g_m(\omega)} \right|^2$$

This factor may then be taken outside the integrals of (13) and there follows from (13) and (14) the result (10):

$$s_g(\omega) = \left| \frac{Y_1(\omega)}{g_m(\omega)} \right|^2 \cdot s_a(\omega) \qquad (10)$$

In any other situation we can combine (13) and (14) only by writing, by analogy with (10),

$$s_g(\omega) = G^2 \left| \frac{Y_1(\omega)}{g_m(\omega)} \right|^2 \cdot s_a(\omega) \qquad (15)$$

where the factor G^2, which may be greater or less than unity, takes into account the 'distributed correlation' between anode and grid currents, as expressed in the variation over λ and z (initial energy and point of emission of the electron) of the factor

$$\left| \frac{Y_1(\omega,\lambda,z)}{g_m(\omega,\lambda,z)} \right|^2 \qquad (16)$$

In many cases of practical interest G^2 will not be much different from unity, since the quotient (16) will tend to be independent of λ at least, but this is by no means necessary.

www.americanradiohistory.com

4. Microscopic Theory

There have already been made [3, 4, 5] various attempts at estimating the noise induced at a control grid in terms of the transit time in an equivalent simple system, such as the 'three-halves power-law.' Unfortunately, however, these treatments have so far made such sweeping assumptions as almost to invalidate subsequent conclusions. For instance, all considerations of the potential minimum and of essential phenomena connected therewith are omitted—it is assumed that all electrons are emitted with the same (zero) velocity, and thus necessarily that all induced electrode currents are similar and are similarly affected by space-charge damping phenomena. This turns out on examination to constitute a serious limitation, for in practice both the induced grid current and the magnitude of space-charge damping effects[6] may vary considerably with emission energy. Further, no account is taken of the fact that the grid is actually at a potential very different from that effective at the grid plane, magnitudes of induced grid currents being dependent on actual grid-cathode potential. Thus in general, no great confidence can be placed in results so far available.

An attempt is being made to develop an approximate 'microscopic' theory of induced grid noise, bearing these points in mind, by considering the effect on the grid of the transit of a particular electron group, and integrating to obtain the accumulated effect. In this article, however, there will be adopted a result of Campbell and others[3] derived by a similar procedure but based on pure three-halves power-law conditions; viz.,

$$s_g(\omega) = \left(\frac{\omega\tau}{5}\right)^2 \cdot 2eI_a\Gamma^2 \quad . . \quad . . \quad . . \quad (17)$$

Here τ is the electron transit time in a three-halves power-law diode, with cathode at the potential minimum and anode at the grid plane, $\omega\tau$ the corresponding 'transit angle' and $2eI_a\Gamma^2$ the anode shot-noise spectrum of the triode.

All of the quantities on the right-hand side of (17) are measurable on the valve, yielding an estimate of the induced grid noise which can immediately be compared with measured values.

5. Measurements.—Neutralization of Lead-Inductance Noise

In making measurements on the induced grid noise, it is convenient to reproduce as closely as possible the assumed conditions of Section 2—that the anode and cathode of the triode are connected to a common earth point by leads of very low impedance. At the frequencies involved, however, there remain inevitably residual anode- and cathode-lead inductance effects which make undesirable contributions to fluctuations observed at the grid. The active component of grid input admittance consists of lead-inductance and transit-time admittances respectively:

$$Y_1 = Y_L + Y_T$$

the total fluctuation at the grid being given, therefore, by

$$s_g(\omega) = \left|\frac{Y_L + Y_T}{g_m}\right|^2 s_a(\omega) \quad . . \quad . . \quad (18)$$

(In this Section and the next it is sufficient to assume that the 'admittance relation' applies in its simple form (10) to the induced effect, as to the lead-inductance effects). The active part of the lead-inductance admittance (see Section 2) is given to a second-order accuracy by

$$Y_L = g_m(\omega^2 L_k C_{gk} - \omega^2 L_a C_{ga}) \quad . . \quad (19)$$

which is purely conductive. The induced component of grid noise is accounted for by the 'transit-time' component of input admittance

$$Y_T = G_T + jB_T \quad . . \quad . . \quad . . \quad (20)$$

and may therefore be expressed as

$$s_{gT}(\omega) = s_a(\omega)\left\{\frac{G_T^2 + B_T^2}{g_m^2}\right\} \quad . . \quad . . \quad (21)$$

In the working range of a valve functioning as a low-noise amplifier, the approximate 3/2 power-law theory suggests, and measurements confirm, that

$$B_T^2 \gg G_T^2 \quad . . \quad . . \quad . . \quad . . \quad (22)$$

Hence to a sufficiently close approximation, one has for the two components of input admittance

$$Y_L = G_L, \qquad Y_T = jB_T \quad . . \quad . . \quad (23)$$

where $B_T = \omega C_t$ is the 'space-charge component' of input susceptance. The total grid-noise spectrum according to (18) is then

$$s_g(\omega) = \left\{\frac{G_L^2 + B_T^2}{g_m^2}\right\} s_a(\omega) \quad . . \quad . . \quad (24)$$

and the induced grid noise, from (21) is

$$s_{gT}(\omega) = \frac{B_T^2}{g_m^2} s_a(\omega) \quad . . \quad . . \quad . . \quad (25)$$

which is obtained from (24) by setting $G_L = 0$. This is in turn effected by satisfying in (19) the 'bridge relation'

$$L_a C_{ga} = L_k C_{gk}$$

by introducing extra reactance into the anode lead and *adjusting this for minimum noise output* from the grid; i.e., from (24)

$$\frac{d}{dG_L} s_g(\omega) = \frac{2G_L}{g_m^2} s_a(\omega) = 0$$

therefore, $G_L = 0 \quad . . \quad . . \quad . . \quad (26)$

and $$s_g(\omega) = \frac{B_T^2}{g_m^2} \cdot s_a(\omega) \quad . . \quad . . \quad . . \quad (27)$$

as required.

It is of interest to note, by way of a check, that substituting for G_L from (19), the mean-square

www.americanradiohistory.com

noise current transferred to the grid circuit from the anode circuit by lead inductance effects alone is given by

$$s_{gL}(\omega) = |\omega^2 L_k C_{gk} - \omega^2 L_a C_{ga}|^2 s_a(\omega)$$

which is the result derived by direct application of circuit theory to the problem.

In order to avoid major difficulties with lead-inductance effects, actual measurements of grid noise were performed on u.h.f. disc-seal triodes, having very low residual lead inductances.

6. Measurement Process

The preliminary sketch given in Section 5 above, while indicating the principles of neutralizing to be observed in dealing with the lead-inductance feedback noise, omits all reference to the effects of associated impedance changes on the measuring instrument. The first stage of the amplifier used for the measurements employs a microwave triode in common grid connection, an equivalent input circuit appearing in Fig. 2.

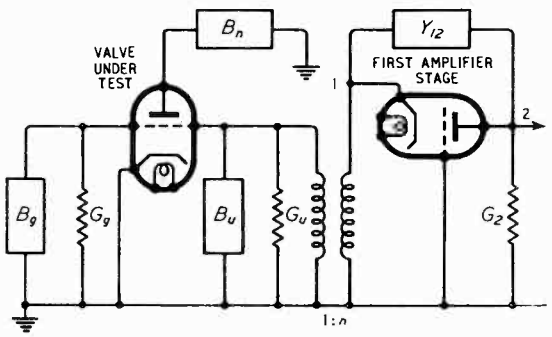


Fig. 2. *Equivalent measurement circuit.*

If $s_g(\omega)$ be the power spectrum of the noise current appearing at the grid of the valve under test, the spectrum of the output from the first stage is given by

$$s_2(\omega) = \left\{\frac{s_g + 4kTG_u}{n^2}\right\}\left|\frac{\Delta_{12}}{\Delta_{22}}\right|^2 + 2eI\Gamma^2\left|\frac{\Delta_{22} - \Delta_{12}}{\Delta_{22}}\right|^2 + 4kTG_2 \quad \ldots \quad (28)$$

In this expression, 1 refers to cathode and 2 to anode of the common-grid amplifier,

$$\Delta_{22} = g_m + g_a + Y_{10} + Y_{12}$$
$$\Delta_{12} = g_m + g_a + Y_{12}$$
$$B_1 = B_g + B_u$$
$$G_1 = G_g + G_u$$
$$Y_{10} = \frac{G_1 + jB_1}{n^2}$$

Where B_g is the grid input susceptance of the test valve and B_u is the adjustable grid circuit susceptance, G_g and G_u corresponding conductances, and n a transformation ratio defined later. g_m and g_a are the mutual and anode conductances respectively and $2eI\Gamma^2$ the anode shot-effect spectrum of the common-grid amplifier. The quantities Y_{12}, the admittance of the feedback path across the amplifier, and G_2, its anode-load conductance, may be neglected here, when the input noise power spectrum equivalent to (28) above appears as

$$s_{1eg} = s_g + 4kTG_u + 2eI\Gamma^2\left\{\frac{G_1^2 + B_1^2}{n^2 g_c^2}\right\} \quad (29)$$

where $g_c^2 = (g_m + g_a)^2$

The procedure in measurement is to note the quiescent value of the output noise associated with s_{1eg} [i.e., corresponding to the value of (29) with $s_g = 0$]; the test valve is then switched on, with the following effects. The input circuit is detuned by the extra space-charge capacitance thrown across it, G_1 is slightly varied by the onset of transit-time and lead-inductance damping G_g at the test grid, and a small fluctuation current is injected by the valve into the new impedance. The capacitance change is tuned out, the conductance change partly neutralized (as will appear later) and the net change in output noise noted. The test valve is then turned off, a standard source (noise diode) tuned on, the circuits retuned and the standard injected fluctuations adjusted to cause the same small deflection of the output meter. The amount of the standard fluctuations is noted, and is assumed to be a measure of the test-valve grid fluctuations.

It is to be noted that s_g, which we wish to measure, is very much smaller than the rest of the terms on the right-hand side of (29). Thus apart from various rudimentary difficulties, the accuracy of the measurement is particularly sensitive to variations during the measurement process of these 'quiescent' terms, due to slight changes of amplifier gain and to variations of the total input-circuit conductance G_1.

The same consideration necessitates also that the amplifier be worked with maximum attainable signal/noise sensitivity, which entails that the impedance viewed from the cathode input terminal shall have a fixed optimum value, in practice very nearly conductive at the mid-band frequency. This optimum conductance is obtained experimentally by adjustment of the transformation ratio $1 : n$ from the grid of the test valve to the input of the amplifier, the value of which will vary according to the value of grid-circuit conductance G_1.

The susceptance B_u is adjusted throughout the measurement independently of any other adjustment to give minimum noise output from

www.americanradiohistory.com

the amplifier; i.e.,

$$\frac{d}{dB_u} s_{1\,eq} = \frac{d}{dB_1} s_{1\,eq} = 2eI\,\Gamma^2 \cdot \frac{2B_1}{n^2\,g_c{}^2} = 0$$

whence $B_1 = 0$ .. .. (30)

It is clear from the work of Section 5 that, if we are to have a reasonable estimate of the induced grid noise (the admittance change associated with it is mainly susceptive and can, therefore, be tuned out) conductance changes G_g at the grid involved in switching on the test valve will have to be small compared say with the resistive grid-circuit conductance G_u. In calculating their effects a first-order accuracy is obtained by assuming that they are, in fact, of differential order compared with G_u. Substituting (30) in (29) we have then that the increase in 'output' noise on switching on the test valve is given in terms of the actual injected noise s_g and the conductance change $dG_1 = G_g$ by

$$s_1 = s_g + 2eI\,\Gamma^2 \cdot \frac{2G_1}{n^2 g_c} \cdot G_g \quad .. \quad .. \quad (31)$$

where by (26)

$$s_g(\omega) = \left\{ \frac{(G_L + G_g)^2 + B_T{}^2}{S^2} \right\} s_a(\omega)$$

S is the test-valve mutual conductance, $s_a(\omega)$ its anode shot effect spectrum and

$$G_g = G_L + G_T$$

the sum of the lead-inductance and transit-time conductances (see Section 5). It is of interest to note at this point that the spectrum of the anode noise in the test valve can be written approximately[6]

$$s_a(\omega) \approx 3 \times 4kT \cdot S$$

the cathode being run at normal temperatures. One has also, for the first amplifying stage

$$2eI\,\Gamma^2 \approx 3 \times 4kT \cdot g_c$$

If then we suppose that by some method we have set G_L (and thus the lead-inductance grid noise) equal to zero, we have from (31) that

$$s_{1\,eq} = 3 \cdot \frac{B_T{}^2 + G_T{}^2}{S^2} \cdot 4kTS + 3 \times 4kT\,g_c \cdot \frac{2G_1 G_T}{n^2 g_c{}^2}$$

i.e.,

$$s_{1\,eq} = 3 \times 4kTS \left\{ \frac{B_T{}^2 + G_T{}^2}{S^2} + \frac{G_T}{S} \cdot \frac{2G_1}{n^2 g_c} \right\} \quad (32)$$

The first term in the bracket of (32) represents the induced grid noise in the valve under test; the second indicates an error in the measurement due to transit-time damping of the input circuit by the test valve. The expression G_1/n^2 is the admittance of the input circuit as viewed from the amplifier, which is fixed by the noise-factor requirements. If the performance of the first valve is of a high standard (i.e., the value of this optimum source admittance is low) the coefficient $2G_1/n^2 g_c$ of the error term may be quite small. It is clear from (29) that there is nothing to be gained by heavy damping of the input circuit (i.e., increasing G_u and G_1) since this will have to be taken up by an adjustment of the transformation ratio n, and will only make the measurements more difficult by decreasing the sensitivity of the measuring gear.

If, as in the elementary neutralizing process discussed previously, we decide to vary G_L until the noise output is a minimum, (31) becomes

$$s_{1\,eq} = 3 \times 4kTS \left\{ \frac{B_T{}^2 + G_L{}^2}{S^2} + \frac{2G_1}{n^2 g_c} \cdot \frac{G_L + G_T}{S} \right\}$$
.. .. (33)

Setting $\frac{d}{dG_L} s_{1\,eq} = 0$

gives $\frac{G_L}{S} = -\frac{G_1}{n^2 g_c}$

and (33) becomes

$$s_{1\,eq} = 3 \times 4kTS \left\{ \frac{B_T{}^2}{S^2} + \frac{2G_1 G_T}{n^2 g_c S} - \frac{G_1{}^2}{n^4 g_c{}^2} \right\} \quad (34)$$

The first term in (34) represents quite closely the induced grid noise, the second term is again the positive transit-time damping error, and the third represents a fixed negative error, depending with the G_T error on the conductance ratio $\frac{G_1}{n^2 g_c}$. The two errors may be expected to cancel somewhere in the working range of the valve. This 'minimizing' process is that most conveniently used in practice.

The above analysis is given in order to demonstrate the errors inherent in measurements of small amounts of induced grid noise, and is not applicable to quantitative correction of the results, since for instance it has not taken into account the effects of impedance changes in the first stage on the feedback conditions in the second. These effects are not, in practice, negligible compared with those we have already discussed, but are not calculable with any degree of confidence.

7. Subsidiary Measurements

Substituting in (25) for the quantity B_T, one has that the induced grid noise spectrum is given by

$$s_g(\omega) = \left| \frac{j\omega C_l}{g_m} \right|^2 \cdot G^2 \cdot s_a(\omega) \quad .. \quad .. \quad (35)$$

where the anode shot noise $s_a(\omega)$ is given at low frequencies by the familiar expression

$$s_a = 2e\,I_a\,\Gamma^2 \quad .. \quad .. \quad .. \quad .. \quad (36)$$

www.americanradiohistory.com

To check (35) against grid-noise measurements, we have to measure or to estimate the behaviour with variable electrode potentials of each of the quantities C_t, G^2, g_m, and s_a on the right-hand side.

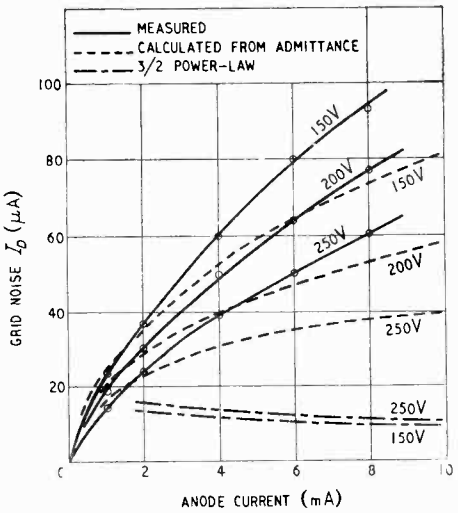


Fig. 3. *Induced grid noise in experimental type E1600, No. 321A, at 200 Mc/s; voltages indicated are those on the anode.*

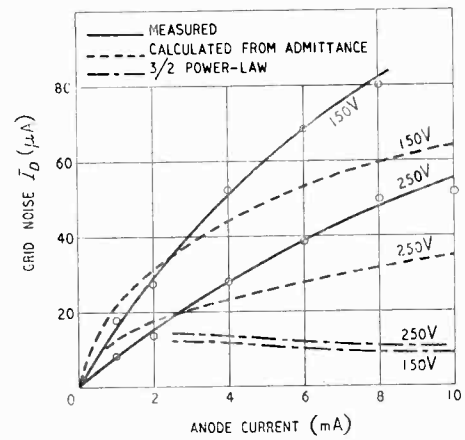


Fig. 4. *Induced grid noise in experimental type E1600, No. 273A, at 200 Mc/s; voltages indicated are those on the anode.*

The capacitance C_t is the space-charge component of grid input capacitance, and is measured as the increase in input capacitance of the valve from cut off to specified working conditions. These measurements, performed at a frequency of 1 Mc/s, need not be discussed further, since a report of very similar measurements has already been published.[7] The mutual conductance was measured under the given working conditions at a low frequency. It is known that neither the capacitance change nor the magnitude of the mutual conductance varies appreciably up to quite high frequencies.

For convenience, the anode shot-noise density (36) was measured at 45 Mc/s. Though this is known to increase considerably at extremely high frequencies, it is not to be expected that there will be appreciable error in assuming that the measured values hold up to 200 Mc/s, at which frequency the grid-noise measurements were performed.

The measured values of mutual conductance g_m and of anode current I_a were also used to determine the equivalent 3/2 power-law transit angle $\omega\tau$ involved in the estimate of grid noise of Section 4. The cathode/grid transit time τ is given, from 3/2 power-law theory for plane triodes[3], as

$$\tau^4 = 5.34 \times 10^{-41}\, \mu_g\, A^2\, (g_m\, I_a)^{-1}$$

where μ_g approximates to the value 1 for a fine-mesh grid, and A is the electrode area.

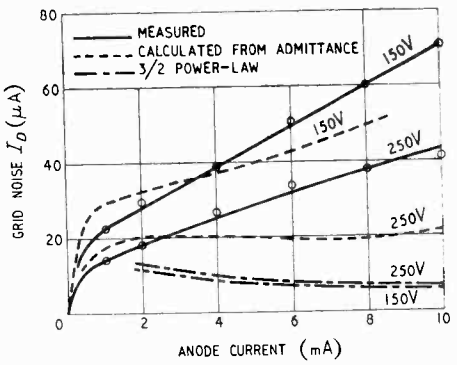


Fig. 5. *Induced grid noise in DET 22, No. 3136, at 200 Mc/s; voltages indicated are those on the anode.*

We have no accurate method so far of determining the value or behaviour with working conditions of the factor G^2 in (35). It is possible, by making various highly arbitrary assumptions and approximations, to deduce from measurements of 'total emission' induced grid noise[8] and an associated admittance change, a qualitative estimate of the behaviour of G^2 with anode current and voltage. It would appear from these measurements that G^2 is, in fact, scarcely influenced by changing electrode potentials, but this is by no means certain.

8. Results

In Figs. 3–6 values of induced grid noise measured on typical high-slope microwave triodes are compared with values predicted from the admittance relation (10) and estimated from the three-halves power-law relation (17). It is assumed that the power spectrum of the fluctuations may be expressed in the form

$$s(\omega) = 2e\, I_D$$

www.americanradiohistory.com

results being given in terms of the pure shot noise in an equivalent temperature-limited current I_D.

It is to be seen that for two E1600 valves measured there is a striking similarity between the functional dependence on valve parameters

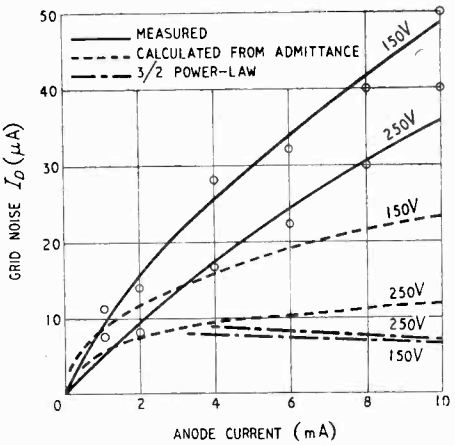


Fig. 6. *Induced grid noise in DET 22, No. 3135, at 200 Mc/s: voltages indicated are those on the anode.*

(anode current and voltage) of the measured and admittance-estimated values of induced grid noise. The 3/2 power-law estimation on the other hand is seen to bear little resemblance to observable fluctuations in any respect.

The DET 22 type of triode, represented in Figs. 5 and 6, has a very much coarser grid than the experimental type E 1600 (grid-pitch to clearance greater) and cannot be expected to follow closely any of the assumptions of internal uniformity which have so far been found necessary in the theory. Considered in this light and bearing in mind the possibilities of considerable experimental error as discussed in Section 6, features of agreement between the predicted and measured curves of Figs. 5 and 6 appear all the more significant.

Reconsideration of Section 6 and, in particular, of Equ. (34) shows that experimental error is qualitatively of the same nature as the discrepancies observable between measured and admittance-predicted grid noise, and may be expected to be large. Thus it is not inconceivable that the predicted results are in fact more reliable as to magnitudes than the measured values themselves.

9. Acknowledgments

The work leading to this report was completed by the author as a student of the University of Durham working at the Research Laboratories of The General Electric Company Ltd., Wembley. He has therefore to acknowledge his indebtedness both to the Manager of the Research Laboratories and to the Board of the Faculty of Applied Science, Durham University, for the extension of invaluable facilities.

REFERENCES

[1] N. R. Campbell, V. J. Francis, E. G. James, *Wireless Engineer*, 1945, Vol. 22, p. 333.
[2] R. L. Bell, *Wireless Engineer*, 1948, Vol. 25, p. 294.
[3] N. R. Campbell, V. J. Francis, E. G. James, *Wireless Engineer*, 1948, Vol. 25, p. 148.
[4] C. J. Bakker, *Physica*, 1941, Vol. 8, p. 23.
[5] D. O. North, *Proc. Inst. Radio Engrs*, 1941, Vol. 29, p. 52.
[6] D. O. North, W. A. Harris, B. J. Thompson, *R.C.A. Review*, 1939, Vol. 4, p. 269.
[7] B. L. Humphreys, E. G. James, *Wireless Engineer*, 1949, Vol. 26, p. 26.
[8] A. van der Ziel, A. Versnel, *Philips Research Reports*, 1948, Vol. 3, No. 1, R.67.

APPENDIX

Induced Grid Noise and Noise Factor

It is interesting to note that the 'distributed' nature of the correlation between induced grid currents and anode shot-noise currents has an effect identical with that of partly *uncorrelated* grid currents (such as the so-called 'total emission grid noise') as far as the operation of the valve as an amplifier is concerned. This may be shown roughly as follows, taking the simple case of the common-cathode triode. Suppose our total cathode-anode shot effect may be represented by

$$s_a(\omega) = \sum_n |i_n(\omega)|^2 \quad \ldots \quad (1)$$

where the $i_n(\omega)$ are elementary shot-effects currents due to the separate electron groups. Then according to the treatment above, the induced noise observable at the grid is of the form

$$s_{ak}(\omega) = \sum_n |\rho_n(\omega) \cdot i_n(\omega)|^2 \quad \ldots \quad (2)$$

If there is a current gain G between grid and anode, the net noise spectrum at the anode terminal is

$$s_a = \sum_n |i_n + G\rho_n i_n|^2 \quad \ldots \quad (3)$$

$$= \sum_n |i_n|^2 |1 + (\rho_{nr} + j\rho_{ni})(G_r + jG_i)|^2 \quad \ldots \quad (4)$$

which on expanding and collecting terms becomes

$$s_a = \sum_n |i_n^2| \{1 + 2\rho_{nr} G_r - 2\rho_{ni} G_i + |\rho_n G^2|\} \quad (5)$$

By proper choice of G_r and G_i this may clearly be made less than the value

$$\sum_n |i_n^2| \{1 + |\rho_n G|^2\} \quad \ldots \quad (6)$$

pertaining to a totally uncorrelated grid fluctuation of the above magnitude

$$\sum_n |i_n \rho_n|^2$$

Considering now the hypothetical case where some of the grid currents are *uniquely* correlated with the anode currents (ρ independent of n) and there exist other grid currents $\alpha^2 \sum_n |i_n|^2$ uncorrelated with anode currents. We have then an output noise spectrum of the form

$$s_a = |1 + \rho G|^2 \sum_n |i_n^2| + \alpha^2 |G^2| \sum |i_n^2| \quad \ldots \quad (7)$$

which by analogy with (5) can be written

$$s_a = \{1 + 2\rho_r G_r - 2\rho_i G_i + |\rho G|^2\} \sum |i_n^2| + \alpha^2 |G^2| \sum |i_n^2|$$

The conditions for minimum noise output on varying the real and imaginary parts of the gain G are

$$2\rho_r + 2G_r |\rho|^2 + 2\alpha^2 G_r = 0$$

i.e., $$G_r = \frac{-\rho_r}{|\rho^2| + \alpha^2} \quad \ldots \quad (8)$$

and $$-2\rho_i + 2G_i \{|\rho^2| + \alpha^2\} = 0$$

www.americanradiohistory.com

i.e., $$G_i = \frac{\rho_i}{|\rho^2| + \alpha^2} \quad \ldots \quad (9)$$

Substituting in (7), the minimum value of noise output due to the valve is given by

$$s_a = \frac{\alpha^2}{\alpha^2 + |\rho^2|} \sum_n |i_n^2| \quad \ldots \quad (10)$$

This minimum has the value zero if the *independent* noise source at the grid is zero; i.e., if $\alpha^2 = 0$.

In general, minimizing (5) with respect to G_r and G_i, one has

$$\sum_n |i_n^2| \{2\rho_{nr} + 2G_r |\rho_n|^2\} = 0$$

i.e., $$G_r = \frac{\sum \rho_{nr} |i_n^2|}{\sum |\rho_n i_n|^2} \quad \ldots \quad (11)$$

Similarly, $$G_i = -\frac{\sum \rho_{ni} |i_n^2|}{\sum |\rho_n i_n|^2}$$

Substitution of (11) and (12) in (4) will clearly give a value for the output noise which is less than that given by (6), but this value *will not in general be zero.*

Thus we see that, for the two models to give identical results, the assumption of a simple unique correlation between anode and grid currents implies also the assumption of an independent internal source of noise at the grid. It is clear that noise phenomena which require the invocation of uncorrelated 'total emission' components of grid noise are, in fact, more accurately explicable in terms of a continuous amplitude- and phase-distributed, but nevertheless complete, correlation between the currents.

NEW BOOKS

Electric and Magnetic Fields

By S. S. Attwood. Pp. 475 + xi, with 276 illustrations. Chapman and Hall, Ltd., 37 Essex St., Strand, London, W.C.2. Price, 44s.

The author is a professor of electrical engineering at the University of Michigan and this is the third edition of the book, which was first published in 1932. In the preface the author says that "in spite of the time devoted during the first two years to the sciences which are basic for electrical engineering, the student needs further training in the development of the fundamental concepts, formulas, terminology, and units used in electric and magnetic field study." This book is intended to provide this training. It is divided into four parts, the first dealing with electric fields, the second with magnetic fields of constant permeability, the third with ferromagnetic fields, and the fourth of only 34 pages with combined electric and magnetic fields; that is, with electromagnetic waves.

Although some parts of the magnetic section have been rewritten to emphasize currents rather than poles, "the magnetic pole concept has not been abandoned entirely since its value as a pedagogical tool is too well established." Considerable attention has been given to field mapping and many diagrams are given of electric and magnetic fields showing the lines of flux and the equipotential surfaces. "A by no means unimportant benefit that should accrue to the student from the presence of many carefully developed field drawings is the stimulation of his imaginative sense." The rationalized m.k.s. system of units has been adopted but tables are given showing the relations between the units in the various systems.

There is a very strange arrangement of material at the beginning of the book. The first eight pages deal with units and dimensions, and if one then jumps to page 23, one proceeds in a normal manner with "elementary facts and definitions," but pages 9 to 22 are devoted to such things as the electronic energy levels in the hydrogen atom, the Lyman, Balmer and Paschen series, and Maxwellian velocity distribution curves. It is difficult to see why this strange material was inserted at this point between the definitions of the international ohm and ampere on p. 8 and Coulomb's law on p. 24. There are other things about the book that are unsatisfactory, and it is surprising to find them in a third edition. In the diagram showing the dependence of dielectric constant and losses on frequency, the dielectric constant is shown going down to 0.5 at certain frequencies, and a curve is stated to give the loss per cycle although the ordinates are labelled power; even when the loss per cycle is shown as constant over a wide range of frequencies, the total loss is also shown as constant and not increasing with the frequency. All this should be carefully explained or omitted.

On looking into the calculation of the force between two parallel conductors and their inductance, one finds the same unsatisfactory treatment. The wire radius is assumed to be small, and the formula for the force is stated to be for $R \ll D$, whereas it applies for any size of wire and the assumption is quite unnecessary and misleading, as is obvious from the fact that the flux density in which one wire is situated, and therefore the force on it, is independent of the size of the other wire. The calculation of the inductance is quite amusing; having calculated the inductance of one wire due to its external flux, the author says "When the flux inside the wire is considered it is difficult to determine the number of flux linkages, so we resort to the energy definition." Here again the mistake is made of assuming that the formula is only correct when $R \ll D$, whereas, if the current density is uniform, the formula is quite correct whatever the size of the conductors, as can be easily shown. It is simpler to assume that the conductors have the same permeability as the surrounding space; any marked departure from this is so rare that it need not be considered in such a book.

The book concludes with a very interesting 8-page historical outline of electricity and magnetism from the earliest times down to the present day.

G. W. O. H.

The Principles of Television Reception

By A. W. Keen, M.I.R.E., A.M.Brit.I.R.E. Pp. 319 + xv. Sir Isaac Pitman & Sons Ltd., Parker St., Kingsway, London, W.C.2. Price 30s.

An unusual feature of this book is the extent to which the American television standards are treated. Throughout, there is a parallel treatment of circuit detail from both the British and the American viewpoints. While this is helpful in showing very clearly how the different transmitting standards affect receiver design, it does make it rather more difficult for those interested primarily in only one system to obtain a comprehensive picture of that system.

The book opens with a table of British and American technical terms. This is a good point, for it emphasizes the differences of language which are liable to be confusing—especially the American use of 'frame' to mean picture. The first chapter describes in simple language the basic principles of television including the scanning-

www.americanradiohistory.com

process and the British and American waveform standards.

The next chapter covers a wide field, for not only does it deal with 'waveforms' and the circuit response to them but the basic r.f. amplifier circuits, including the earthed-grid stage, are also treated. As one might expect there is no analysis of the circuits but merely an elementary description of how they work. The transient response section, however, is quite well done and gives a very good account of the effect of simple circuits upon pulses of various kinds.

Chapter III covers the c.r. tube with its focusing and deflecting arrangements and, again, a good general picture of the subject is given without going into it very deeply. E.h.t. supplies are also discussed. Chapter IV is devoted to time bases and covers, not only the saw-tooth and pulse voltage generators, but current amplifiers for magnetic deflection and linearizing circuits. A chapter on sync separation follows.

A single chapter covers the whole of vision-signal amplification in only 40 pages. As v.f., i.f. and r.f. amplifiers, detectors, gain control, sound-channel rejection, frequency-changers and noise are covered, it is plain that the treatment can only be superficial.

The sound channel has a chapter to itself, but is not much more detailed since, in addition to the British a.m. and the American f.m. systems, methods using pulse-width modulation are also described.

There is a lengthy and well-illustrated chapter on the complete receiver in which circuit diagrams are given of several British and American sets. Aerials and feeders have a short chapter, receiver test equipment another and one on colour television ends the book proper. There is a short appendix containing a number of useful mathematical expressions. Each chapter contains a useful, if condensed, bibliography.

In his preface, the author states that the book has grown out of material originally produced as a short correspondence course for engineers of a receiver manufacturer's servicing organization. It is certainly not a servicing book and is not intended to be. It is meant to cover principles only and it does so. Whether or not it does so adequately is another matter and a debatable one, since it depends on how much one expects a service technician to know. In the reviewer's opinion it does no more than lay a foundation for further study.

Mathematics are almost non-existent in the body of the book and even in the Appendix are quite simple. The book can be recommended for those who have little or no knowledge of television but who have a good background of ordinary wireless theory and practice. It will enable them to obtain a good general, but hardly a detailed, knowledge of television. The book is not entirely free from errors, but where they are not obvious they are unlikely seriously to hamper the student.

W. T. C.

Quirlende elektrische Felder

By Fritz Emde. Pp. 118 + vii, with 41 illustrations. Friedr. Vieweg und Sohn, Braunschweig, Germany.

Dr. Emde, who was formerly Professor of Electrical Engineering at Stuttgart, is known to a wide circle outside the field of electrical engineering through his Tables of Mathematical Functions. This present small volume on electric fields was handed over to the publishers in 1943 but has only just been published. It is confined to whirling or curl fields, that is, to the relatively small fields associated with magnetic fields, and does not deal with the powerful electric fields occurring in high voltage work. The author states that his main object is to give the reader a clear conception of the principles underlying the fundamental laws of Maxwell, Heaviside, etc.

The treatment is very full and clear and illustrated by simple numerical examples. On p. 48 a field is resolved into its two components, one a curl field and the other a curl-free field; this is done very simply for the case of a packet of iron laminations with numerical examples. Several well-known paradoxes, such as the revolving bar magnet and Hering's spring tongs, are examined on p. 72. Several pages are devoted to the movement of dielectrics in a magnetic field, and Wilson's classical experiments are described. The final twenty pages are more mathematical, dealing with divergence, rotors, Spreizung, Drillung and Affinors, but in the introduction Dr. Emde says that he trusts that those readers who put the book down without working through it to the very end will have gained something from it. The author decided to omit electromagnetic waves as they are treated so fully in a number of books. This is a book that can be unreservedly recommended to anyone with a knowledge of German.

G. W. O. H.

Ein Ultrakurzwellen-Telefoniesystem hoher Kanalzahl mit Frequenzweiche

By G. C. Fontanellaz. Pp. 74, with 41 illustrations. Verlag Leeman, Zürich. Price 10 francs (Swiss).

This is the author's doctorate thesis at the Zurich Hochschule, on work carried out under the guidance of Dr. Tank.

The first 30 pages are devoted to a discussion of the general problem of multiple-channel telephony, with a very large number of channels. The remainder of the book contains a description of a proposed system which was constructed and tested in the laboratory. Photographs of the actual apparatus are given. The system contains 6 u.h.f. bands, each with 12 speech channels; the aerial is of the highly-directive narrow-beam type. The frequencies discussed are between 200 and 3000 Mc/s. Various methods of producing and modulating the pulses are discussed. It should be noted, however, that this is the fourth monograph published by the Hochschule during 1949 dealing with closely interlinked subjects, or branches of the same subject, and anyone interested in the subject would be well advised to obtain the four monographs. The others were "Ultrakurzwellen Frequenzweiche," by Staub (7.50 f), "Breitband Richtstrahl Antennen," by Peter (8 f), and "Impulsmodulation," by Bachmann (9 f).

G. W. O. H.

Stetige Regelvorgänge (Servo-mechanisms)

By Winfried Oppelt. Pp. 144, with 42 illustrations. Wissenschaftliche Verlagsanstalt K.G., Hannover. Price DM.7.80

In March 1948, p. 88, we reviewed a book by the same author entitled "Grundgesetze der Regelung" and stated that it contained the basic mathematical equipment, the practical application of which would be dealt with in another volume. This is the volume referred to. It deals with every possible type of servo-mechanism, mechanical, electrical and thermal, with diagrams of the mechanism, and numerous graphs illustrating their operation, stability, etc. The book concludes with a very full bibliography of the subject.

The book was completed, and ready for publication, in 1945; the author states in the preface that subsequent developments are of course, not dealt with in the text, but the bibliography has been brought up to date. The new volume is not an austerity production as was the first; the production, and especially the illustrations, are excellent and the two volumes form a valuable introduction to a subject of growing importance. We commend them to anyone with even a small knowledge of German; so much of the material is graphic.

G. W. O. H.

www.americanradiohistory.com

SIGNAL-TO-THERMAL NOISE RATIO

Comparison of F. M. and A. M. Receivers

By M. V. Callendar, M.A., A.M.I.E.E.

SUMMARY.—Curves are given for output signal/noise ratio for thermal noise with an f.m. receiver, including conditions where the signal is equal to or less than the noise, as well as the well-known case where the signal greatly exceeds the noise. Corresponding curves are also given for an a.m. receiver, and a comparison is drawn between the results for f.m. and a.m., bringing out especially the effect of the ratio of pre-detector to post-detector bandwidth in each case.

It is concluded that, apart from its well-known advantages for high-quality broadcasting, f.m. can give results at least equal to a.m. in respect to signal/noise ratio for communication work provided that the pre-detector bandpass is no wider than necessary to accept the full deviation, and that the limiter is fully efficient: the deviation ratio for such work should not exceed about four. However, these conditions are not always easy to comply with in practice (for example, in cases where a wide bandwidth is necessary from frequency-drift considerations) and thus an a.m. receiver will, in some instances, be found to give a slightly better range of intelligible reception than its f.m. counterpart, particularly if any slight mistuning be present.

1. Introduction

THE question of signal/noise ratio in f.m. receivers has been discussed by many writers since Armstrong's original paper[1]. In that paper, the improvement in signal/noise ratio for an f.m. receiver as compared with an a.m. receiver was evaluated for the case where the signal at the detector is considerably greater than the noise; this analysis is all that is necessary when considering good quality broadcast reception, but sheds little light on the important question of the relative merits of f.m. and a.m. in respect of maximum range of intelligible communication. In a recent paper[2] the present author evaluated formulae for signal/noise ratio in a.m. receivers and examined its variation with bandwidth. Since that paper was written a paper[3] by Stumpers on noise in f.m. receivers has been published; on the basis of his results we can calculate signal/noise curves for f.m. down to below the noise level, and thus compare f.m. and a.m. receivers at any level, and for any bandwidth.

As regards the requirements in audio signal/noise ratio for the various types of service, the following table may be taken as a fair indication:

Telegraphy	0 db
Bare minimum intelligibility (telephony)	10 db
Comfortable intelligibility (telephony)	20 db
Minimum acceptable for broadcast listening	30 db
Good quality broadcast reception	40 db

2. Theory

The results of Stumpers' excellent mathematical paper[3] are expressed in a series of curves for a.f. noise output against noise-to-signal ratio at the input to the detector, for various value of the ratio $\Delta f/f_a$ where $2\Delta f$ is the pre-detector bandwidth and f_a is the audio bandwidth. The detector is here assumed perfect; i.e., it responds

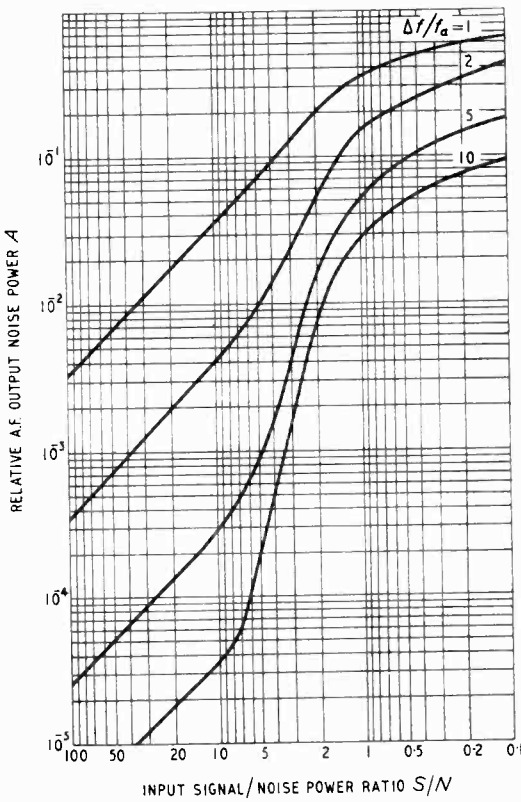


Fig. 1. *Variation of output noise from an ideal f.m. detector with input signal/noise ratio. For a rectangular pre-detector filter of bandwidth $2\Delta f$ and audio band f_a. From Stumpers[3] (Figs. 2 and 3), who gives noise output* $= A\Delta f^2/_2$.

MS accepted by the Editor, July 1949

www.americanradiohistory.com

only to changes in frequency and not to changes in amplitude. Curves for the case of a rectangular pre-detector bandpass characteristic are reproduced here as Fig. 1, and clearly form a basis from which signal/noise ratio can be obtained : in their existing form, however, they are apt to give a misleading impression of the relation of signal/noise ratio to bandwidth since the 'noise-to-signal ratio' used as abscissa is itself a function of bandwidth (see below).

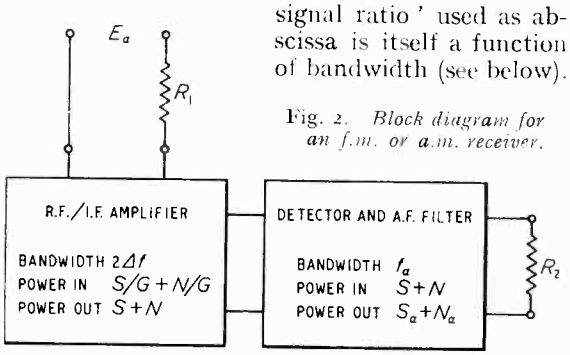


Fig. 2. *Block diagram for an f.m. or a.m. receiver.*

Stumpers also gives a formula for the suppression (due to limiter action) of the signal modulation by the noise when the noise-to-signal ratio is large ; the amplitude of the signal modulation is reduced in the ratio $B = (1 - e^{-S/N})$ where N is noise and S is signal, both in terms of power at the input to the detector.

Assuming that the frequency detector is 'perfect' (i.e., that it includes a perfect limiter as well as the usual frequency to amplitude converter and amplitude detector) we may define a conversion factor C such that the detector gives an output of C watts per cycle deviation at the input. The assumption that C is invariable under all conditions (of input amplitudes, etc.) is made by Stumpers ; this assumption is evidently not true if the input to the detector diodes is too small, but this is perhaps unlikely in practice (since the detector follows the limiter). However, it is necessary to take into consideration the suppression of modulation by noise which must occur in the diodes which 'convert' the i.f. signal (here amplitude modulated) from the discriminator into the required audio signal. This correction factor D is small and only appreciable for very low signal/noise ratios ($S/N <$ say 6 db) where we may assume the output noise spectrum roughly uniform (Stumpers' Fig. 6) : thus we can put $D = 1/(1+N'/2S')$ approx. (vide formula for a.m. below) where N'/S' is the output noise/signal ratio for $\Delta f = f_a$ as calculated direct from Stumpers.

Then signal a.f. output power =

$$S_a = m^2\delta f^2 \times C \times D \times B^2$$

where δf is peak deviation, m is fractional modulation, and B and D are modulation-amplitude suppression-ratio factors for the limiter and the detector respectively. And noise a.f. output power $= N_a = A \times \Delta f^2/2$ multiplied by $2C$, since Stumpers assumes a d.c. output of 1 V across 1 Ω for 1-c/s deviation. Here A is the ordinate in Fig. 1, and $2\Delta f$ is pre-detector power bandwidth. Thus the a.f. signal/noise ratio =

$$\frac{S_a}{N_a} = \frac{m^2\delta f^2}{\Delta f^2} \cdot \frac{B^2 D}{A}$$

If, as in Fig. 2, a signal e.m.f. (E_a) from a source of resistance R_1 is applied to a receiver having a total pre-detector power gain G, the signal power at the input to the detector will be $G \times E_a^2/4R_1$, assuming matching of the receiver to the source. The corresponding noise power will be $G \times FKT \,.\, 2\Delta f$, where F is the receiver 'noise factor' and KT has the usual significance (noise power available from a resistance per cycle bandwidth).

Thus the signal/noise power ratio at the input to the detector $= \dfrac{S}{N} = \dfrac{E_a^2}{8FKT\Delta f R_1}$

For convenience, we will here define a parameter $S_r = \dfrac{S}{N} \cdot \dfrac{\Delta f}{f_a}$ and call this the "relative signal input." Then we have $S_r = \dfrac{E_a^2}{8FKTR_1 f_a}$ and this represents the signal relative to the

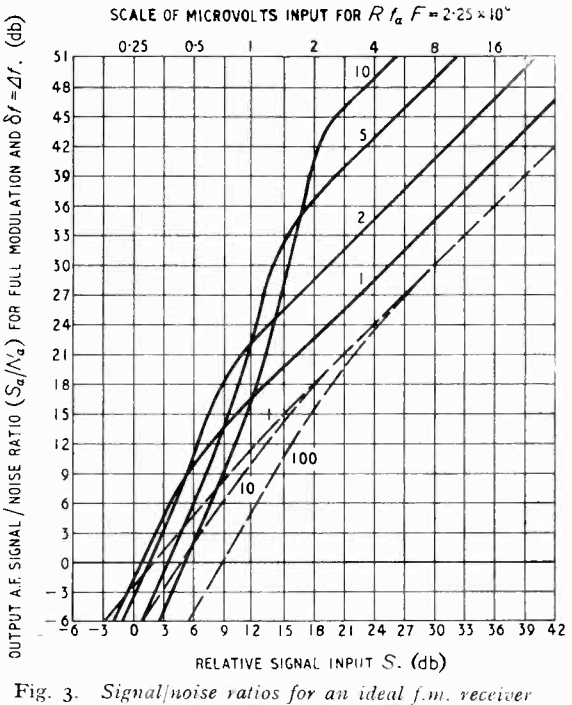


Fig. 3. *Signal/noise ratios for an ideal f.m. receiver compared with those for an a.m. receiver. Number marked on curves is ratio of half pre-detector bandwidth to post-detector bandwidth ($\Delta f/f_a$). Full lines for f.m., dashed lines for a.m.*

www.americanradiohistory.com

noise in a bandwidth $2f_a$. We can now plot curves (Fig. 3) for a.f. signal/noise ratio against S_r, which is independent of the pre-detector bandwidth: for simplicity, we take the case where $m = 1$ and $\Delta f = \delta f$: the actual signal/noise ratio for any lower deviation will be proportionally less.

For comparison, we have also plotted the signal/noise ratio for an a.m. receiver from the relation:

$$\frac{S_a}{N_a} = \frac{m^2 S}{N} \cdot \frac{\Delta f}{f_a} \cdot \frac{1}{(1+N/2S)} = \frac{n_r^2 S_r}{(1+N/2S)}.$$

This is taken from Section 2.5 of the paper[2] by the present writer cited above; it refers to a parabolic detector but holds also for a linear detector to within, say, ± 1 db at all inputs. In the case of a linear detector, the factor $1/(1 + N/2S)$ allows for the suppression of modulation by noise when the signal/noise ratio is low: with a parabolic detector, the same factor merely expresses the quadratic relation between input and output.

The curves as plotted give all the information necessary to compare theoretical signal/noise ratios for a.m. and f.m. for any ratio of pre-detector to post-detector bandwidtn; such comparisons will be unaffected by the a.f. bandwidth, the noise factor, or the signal-source impedance. However, if we require the numerical signal/noise ratio for any particular case, it can readily be re-plotted in terms of microvolts input (E_a) from the equation given above relating S_r to E_a for any given values of F, R_1 and f_a. A scale of microvolts for the particular case of a typical v.h.f. receiver having $R_1 = 75\,\Omega$ $F = 10$ and $f_a = 3$ kc/s, has been added to the figure by way of illustration.

When the signal is much larger than the noise ($S/N > 10$) the curves conform to the simple formula $\frac{S_a}{N_a} = 3 \times \frac{S}{N} \times \frac{\delta f^3}{f_a^3}$ which agrees with Armstrong's original formula, for signal/noise ratio. Fig. 4, derived directly from Fig. 3, gives the effect of increasing the pre-detector bandwidth without altering the deviation, and agrees with the above simple formula in showing that when $S/N <$ about 10, an increase in pre-detector bandwidth does not affect the a.f. signal/noise ratio for a given deviation ratio.

It should be noted that the curves of Fig. 1 assume a rectangular selectivity curve for the pre-detector circuits: Stumpers also gives figures for a curve of Gaussian form, but the results from this are not substantially different and do not, in the present writer's view, approximate any more nearly to practical conditions. In practical cases, where the curve is not rectangular, Δf must be taken as the 'power bandwidth' defined by $2\Delta f = \frac{1}{S_{max}} \cdot \int_{-\infty}^{+\infty} S.\, df$

A further limitation on accuracy is that the noise output is not strictly independent of deviation of the carrier: (see Stumpers[3] p. 1090). It is felt that it is justifiable to neglect this second-order effect in practice, since it will only occur at large deviations where the noise will be usually masked by the modulation. This effect should not be confused with the large effects of deviation or detuning when the limiter is not in operation and the balance of the discriminator is relied upon for noise reduction. However, it does mean that signal/noise ratio deteriorates appreciably with mistuning even with a perfect limiter in action.

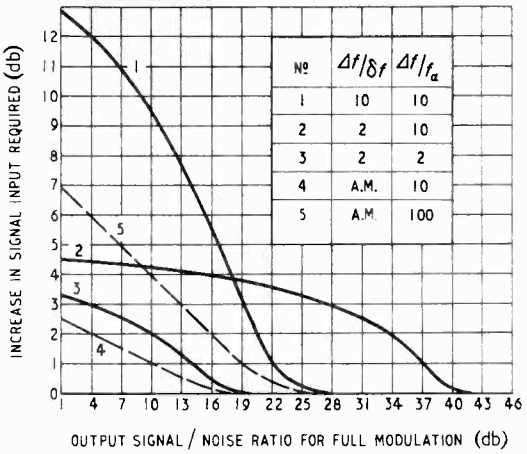

INCREASE IN SIGNAL INPUT REQUIRED (db)
OUTPUT SIGNAL / NOISE RATIO FOR FULL MODULATION (db)
№ | Δf/δf | Δf/f_a
1 | 10 | 10
2 | 2 | 10
3 | 2 | 2
4 | A.M. | 10
5 | A.M. | 100

Fig. 4. *Increase in signal input required to restore output signal/noise ratio if pre-detector bandwidth is made wider than ideally necessary (i.e., if $\Delta f > f$ for f.m. or if $\Delta f > f_a$ for a.m.). Full lines for f.m., dashed lines for a.m.*

Curves are also given by Stumpers for the case where *pre-emphasis* at the transmitter enables a de-emphasis filter to be inserted at the receiver. For the case of a high fidelity system where $f_a = 15$ kc/s and the filter time constant is 75 microseconds (3 db cut at 2 kc/s), the output signal/noise ratio is improved by 12 db approx. when S/N is > 10. For lower values of S/N, the output noise spectrum progressive changes from one increasing linearly with frequency to one roughly uniform over the audio band (and thus similar to that obtained on a.m.): the improvement from pre-emphasis thus becomes smaller and eventually falls to about 6 db for $S/N = 1$ on f.m., where it is equal to that obtained for a.m. For a communication system having $f_a = 3$ kc/s, since the maximum permissible pre-emphasis is about 3 db at 1.5 kc/s (time constant 100 microseconds) it is clear that the gains in signal/noise ratio will be much less and for an output signal/noise ratio of 10 to 20 db (i.e., for S/N values

www.americanradiohistory.com

around 1 to 2) the gain from pre-emphasis will be small (about 2 db) and not materially different for a.m. and for f.m.

3. Practical Conclusions

It is important to emphasize at the outset that, while there is no reason why the curves of Fig. 3, for a.m. should not hold under practical conditions, the curves for f.m. represent an *ideal receiver*, and results from actual receivers will be worse for two reasons :—

(a) It is assumed that the peak deviation (δf) can be made equal to half the power bandwidth. Even for a good 'square' bandpass, this would entail loss of linearity for high modulations and, in practice, a Δf of $1.3\delta f$ to $2\delta f$ must be allowed for. From Fig. 4, an increase of 1 to 4 db in signal input will be required to restore the a.f. signal/noise ratio when the latter is of the order of 10–20 db. In certain cases, notably where a small deviation is used at v.h.f., the i.f. bandwidth will have to be made much greater than the deviation in order to cover possible frequency drift : here the signal/noise ratio will be noticeably worsened.

(b) The detector will not necessarily be completely insensitive to the amplitude modulation energy in the noise. The correction required here will be less than 1 db with really good limiting and accurate tuning, but may be quite large for small signals in receivers of inadequate sensitivity: in such cases, the signal/noise ratio will be sharply dependent upon the exact centering of the carrier and the symmetry of the tuned circuits, and the noise will increase notably with modulation.

An increase of 3 db in signal at small inputs, falling to 0 db at larger inputs (see Fig. 4) is considered adequate to cover the above factors for a really *efficient practical f.m.* receiver : our conclusions for such a receiver are expressed in the table.

TABLE

R.F. INPUTS required for a given output signal/noise ratio (expressed relative to an a.m. receiver having $\Delta f = f_a$.)

	$\Delta f/f_a$	Required A.F. Signal/Noise Ratio			
		0 db	10 db	20 db	40 db
Amplitude Modulation	1	0 db	0 db	0 db	0 db
	10	+3	+1	0	0
	100	+7	+4	+1	+0
Frequency Modulation	1	+2	−2	−4	−4
	2	+3	−2	−9	−10
	5	+5	−1	−8	−18
	10	+6	+2	−6	−22

In this table, comparisons are of course only to be made in the vertical columns : the deviation ratio may be taken as three-quarters of the band ratio quoted in the table.

As regards, then, a communications receiver for telephony, we conclude :—

(i) The f.m. receiver should be equal to, or slightly better than, the a.m. receiver as regards range of intelligible reception, *provided* that it has (a) a limiter fully operative at the smallest inputs, (b) an i.f. curve shape not far from rectangular and an i.f. bandwidth not exceeding 3 times the deviation, (c) a suitable deviation ratio (not exceeding 4), A small improvement (some 2 db), may be obtained for both a.m. and f.m. by pre-emphasis. Accurate tuning is assumed.

(ii) If, however, these conditions are not satisfied, the a.m. receiver will tend to have an advantage. In particular, the effective sensitivity of an a.m. receiver is but little affected in most practical cases by the pre-detector bandwidth (see previous paper[2] for examples) but, in the case of f.m. communication receivers, it is essential that the pre-detector bandwith be kept down to the minimum compatible with the requirements of the signal deviation. As an example, an increase in bandwidth from $\Delta f = 5f_a$ to $\Delta f = 10f_a$ could lower the a.f. signal/noise ratio from 18 db to 6 db (for $m\delta f = 4f_a$ and $S_r = 10$ db). For high quality broadcast reception, however, the pre-detector bandwidth ceases to be important, as indicated by Fig. 4.

As an extreme case where a.m. will give superior signal/noise ratio and range, we may take a miniature portable receiver on, say, 160 Mc/s, where the normal $\pm 0.01\%$ overall frequency stability on both receiver and transmitter imposes a bandwidth of at least 80 kc/s for − 3 db and gain is insufficient (due to space and weight limitations) to permit of efficient limiting below, say, 10 μV. Accurate centering of the receiver discriminator on the transmitter is likewise impractical and it is clear enough that the input for 10 db signal/noise ratio for f.m. will be appreciably greater than that for a.m., whether wide or narrow deviation is used. At the other extreme, of course, we have the usual high quality broadcast receiver, where the range for 40 db audio signal/noise ratio will be nearly three times better with f.m. than with a.m.

In the only practical example in which the writer has made actual comparative measurements (a battery-operated Services f.m. set), it was found that an input of 1.2 μV was the minimum for intelligible signals with f.m. and gave an a.f. signal/noise ratio of 0 db for 30% modulation : the same figure for minimum input was obtained on changing the detector to operate on a.m. In this case, the limiter was not fully operative at the inputs in question and the i.f. curve was that

www.americanradiohistory.com

obtained from four tuned circuits in cascade: Δf was about 15 kc/s and δf was about 10 kc/s.

If the receiver was slightly detuned, results were of course worse on f.m. than on a.m.

Since the above was written, a paper entitled "Theoretical Signal-to-Noise Ratios in F.M. Receivers" by D. Middleton, has appeared in the *Journal of Applied Physics* for April 1949. This paper treats the subject exhaustively from a mathematical point of view; the practical aspects are touched on somewhat briefly, and the conclusions are slightly different from those in the present paper, owing to different assumptions (mainly a Gaussian shape of response characteristic in place of a rectangular shape).

REFERENCES

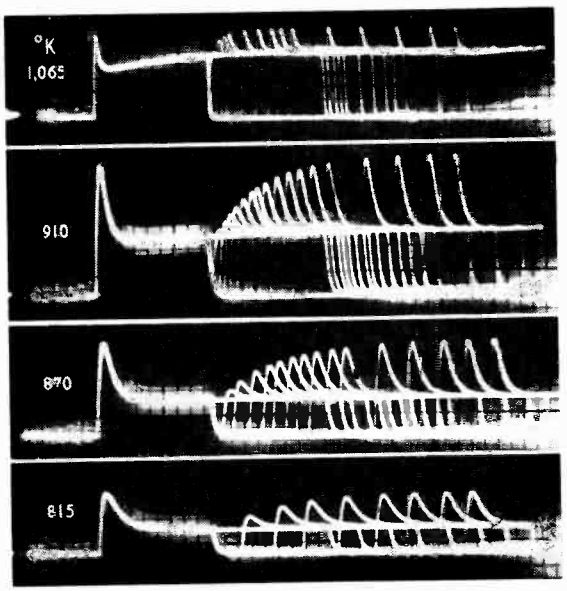
1 Armstrong. "Frequency Modulation," *Proc. Inst. Radio Engrs*, May 1936.
2 Callendar. "Thermal Noise in A.M. Receivers," *Wireless Engineer*, Dec. 1948.
3 Stumpers. "Theory of F.M. Noise," *Proc. Inst. Radio Engrs*, Sept. 1948.

CORRESPONDENCE

Letters to the Editor on technical subjects are always welcome. In publishing such communications the Editors do not necessarily endorse any technical or general statements which they may contain.

Change of Mutual Conductance with Frequency

SIR,—The October 1949 *Wireless Engineer* published a paper[1] on this subject which was based on the operating characteristics of pentode-type valves. New valves exhibited little frequency dependence of the mutual conductance, g_m, but valves operated over 1,000 hours showed marked reduction in g_m with decreasing frequency between 10^7 and 10^5 c/s. This frequency dependence of gain was explained as arising from a parallel resistance-capacitance impedance in the cathode circuit causing negative feedback. This impedance was associated with a change in the physical nature of the oxide cathode interface. The hypothesis was offered that the electrical contact between the oxide coating and the base metal deteriorated during the life of the valve. When this occurred only a few points of contact bridged the thin vacuum gap separating the base metal and coating, thus forming effectively a parallel RC impedance to the normal flow of electron current from the base metal into the coating.



Anode current pulses of 7AD7 pentode with 15-μsec grid pulse driving the grid from — 20V to zero with an anode supply of 400 V.

A study[2] has been made of the decay of anode current passed by a pentode when a square voltage pulse is applied to the grid. These results were also interpreted in terms of a parallel RC impedance at the oxide cathode interface but the physical model offered to explain the impedance used the development of a homogeneous interface compound, sandwiched between the coating and base metal, having a low conductivity. It seems likely that this rather than the inhomogeneous, poor contact hypothesis is the correct explanation of the change of g_m observed by Raudorf. The figure shows the anode current pulses passed by a type 7AD7 pentode upon the application of a 15-μsec grid pulse sufficient to drive the control grid from cut-off to zero bias. The decay is readily apparent and is more pronounced at lower temperatures. Following the initial pulse a second pulse is applied after a varying time delay. When the time delay is short the initial current is less than that for the first pulse. A number of these for different delays are superimposed to form the figure. The recovery of anode current to the initial value obtainable at the beginning of the first pulse may be seen. The decay constant (time for the plate current to decrease to $1/e$ of its initial value) and the recovery constant, similarly defined, are seen to be different and to be temperature dependent. Using the equations for degenerative feedback amplifiers the two-time constants are:

$$\text{Decay constant,} \quad D = RC/(1+g_mR) \quad (1)$$

$$\text{Recovery constant} \quad D' = RC \quad (2)$$

where R and C are the resistance and capacitance of the interface. From an experimental evaluation of D, D' and g_m values of R and C may be obtained. Tabulated below are the results of this study which seem pertinent to the association of the impedance with an interface compound.

1. Cathodes of valves made using a pure nickel base metal do not develop the decay.
2. Cathodes of valves made using a nickel base metal containing 0.1 to 0.2 per cent silicon do develop the decay.
3. X-ray diffraction patterns of cathodes made on silicon nickel show an interface of Ba_2SiO_4 approximately 10^{-4} cm thick. New cathodes have no decay and no interface compound.
4. The measured interface resistance and thickness are consistent with values[3] of the specific electrical conductivity of Ba_2SiO_4. The temperature depend-

www.americanradiohistory.com

ence of resistance is that to be expected of a semiconductor.

5. The measured interface capacitance is independent of temperature and is consistent with the measured interface thickness and an assumed dielectric constant of 10.

Similarities between our observed decay phenomena and the reported frequency dependence of g_m are as follows:

1. Both effects develop during the same life period of the valve.
2. Decay constants of about 1 μsec are consistent with the frequency dependence between 10^7 and 10^5 c/s.
3. Values of R and C obtained using the two methods of analysis are essentially the same, $C \approx 2 \times 10^4$ pF and $R \approx 100\ \Omega$.
4. Cathode sparking has been associated with the development of both interface models, see Ref. 3.

Because of these similarities it is suggested that the two phenomena arise from the same physical model. Since the base metal composition used is not specified by Raudorf it cannot be assumed that the interface compound in his valves is Ba_2SiO_4; however, many reducing impurities found in commercial cathode cores are capable of forming other interface compounds.[4] In addition to adequately accounting for the frequency dependence of g_m our model based on a growth in interface thickness with life offers an explanation for the reduction in capacitance with increasing resistance as reported by Raudorf. This was not explained by the poor contact hypothesis which accounts only for the resistance increasing with cathode life.

From the standpoint of practical cathode operation it is of interest to note that the decay phenomena develops more readily[5] in valves operated under cut-off conditions than under class A operation. This we attribute to the development of unactivated Ba_2SiO_4 in the former case since the interface is formed in both types of operation. Previous study[3] showed that this interface compound could be activated by electron current flow. A detailed account of these and other experiments is being prepared for publication elsewhere.

Department of Physics, A. EISENSTEIN
University of Missouri.

[1] W. Raudorf, *Wireless Engineer*, October 1949.
[2] A. Eisenstein, paper presented at American Physical Society meeting, November 25, 1949.
[3] A. Eisenstein, *J. appl. Phys.* Vol. 20, No. 776, 1949.
[4] D. A. Wright, *Proc. Phys. Soc.* B, Vol. 62, p. 188, 1949.
[5] Results obtained at the M.I.T. Servomechanisms Laboratory. The writer is indebted to E. S. Rich of that laboratory for supplying valves used in this study.

A Problem — An Answer

In the February issue we published an enquiry from a reader. Mr. A. T. Starr, of Standard Telecommunication Laboratories Ltd., gives the following answer:—

"Regarding the problem of finding the monotonic function with the fastest rate of rise, the solution is very old and is contained in almost any book on Analysis or Operational Calculus. It is well known that

$$\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} e^{\lambda t}\, d\lambda/\lambda = 0,\ t < 0.$$
$$= 1,\ t > 0.$$

To prove this it is only necessary to complete the contour by a large semi-circle to the right for $t<0$, when the integral is zero: when $t>0$ the semi-circle must be to the left and then there is a pole at $\lambda = 0$ at which the residue is unity. Thus if $f(\lambda) = 1/\lambda$, the time function rises instantaneously from zero to unity at $t = 0$: it is thus monotonic and has an infinite rate of rise.

If I may guess what is in the reader's mind, I should say that he wants the solution of an amplifier with a monotonic response and fastest rate of rise. This problem cannot be said to be solved, but it is possible that the work of W. E. Thomson, *Wireless Engineer*, January 1947, has approached, if not actually obtained, the solution." [ED.]

STANDARD FREQUENCY TRANSMISSIONS

From 1st February 1950 standard frequencies are being transmitted from Rugby on 60 kc/s, 5 Mc/s and 10 Mc/s with powers of 10 kW. The times of transmission are: 60 kc/s, 1029–1045; 5 Mc/s, 0544–0615; 10 Mc/s, 0629–0700; all times being G.M.T. The accuracy is within two parts in 10^8 and is monitored by the National Physical Laboratory to which all enquiries and comments should be addressed at Teddington, Middlesex.

Each transmission is modulated with the following 15-minute cycle:—

Modulation, Slow morse call sign (MSF) followed by speech, at 59–00, 14–15, 29–30, 44–45, minutes past the hour.
Modulation, 1,000-c/s tone, at 00–05, 15–20, 30–35, 45–50 minutes past the hour.
No modulation, at 05–14, 20–29, 35–44, 50–59 minutes past the hour.

INSTITUTION OF ELECTRICAL ENGINEERS

The following meetings will be held at the I.E.E., Victoria Embankment, Savoy Place, London, W.C.2, commencing at 5.30:—

March 15th, "The Structure, Electrical Properties and Applications of the Barium-Titanate Class of Ceramic Materials," Prof. Willis Jackson.

March 27th, "The Operation and Maintenance of Television Outside-Broadcast Equipment," T. H. Bridgewater.

March 30th, Symposium of Papers on the M.K.S. System of Units.

April 3rd, Discussion meeting on "The Place of the Model in Teaching," opened by D. G. Sandeman.

April 12th, "A Review of some Television Pick-up Tubes," J. D. McGee, and "The Design of a Television Camera Channel for use with the C.P.S. Emitron," E. L. C. White and M. G. Harker.

BRITISH INSTITUTION OF RADIO ENGINEERS

During March the following meetings will be held:—

London Section. March 23rd, at 6.30, at the London School of Hygiene & Tropical Medicine, Keppel St., W.C.1; "High Performance Television Monitors," by J. E. Jacobs.

West Midlands Section. March 22nd, at 7.0, at the Wolverhampton & Staffordshire Technical College, Wulfruna St., Wolverhampton; "Radio Interference with Broadcast Reception," by A. A. Devey.

North-Eastern Section. March 15th, at 6.0, at Neville Hall, Westgate Rd., Newcastle-on-Tyne; "Electrical Breakdown of Gases at Ultra High Frequencies," by Dr. Prowse.

South Midlands Section. March 30th, at 7.0, at the Coventry Technical College, Room A5. "Electronics and the Brain," by H. W. Shipton.

www.americanradiohistory.com

WIRELESS PATENTS

A Summary of Recently Accepted Specifications

The following abstracts are prepared, with the permission of the Controller of H.M. Stationery Office, from Specifications obtainable at the Patent Office, 25, Southampton Buildings, London, W.C.2, price 2/- each.

ACOUSTICS AND AUDIO-FREQUENCY CIRCUITS AND APPARATUS

627 639.—Moving-coil loudspeaker in which the output transformer acts as the field-producing magnet; this is stated to give a volume expansion effect.

E. K. Cole Ltd. and E. L. Hutchings. Application date April 3rd, 1947.

AERIALS AND AERIAL SYSTEMS

627 689.—Aerial for ships for use at u.h.f. employing conductive cones designed to give omni-directional radiation.

Marconi's Wireless Telegraph Co. Ltd. (assignees of J. B. Atwood). Convention date (U.S.A.) May 4th, 1942.

627 091.—Tunable waveguide in which the tuning cup is associated with a loading element adjustable relatively to the cup.

Hazeltine Corporation (assignees of J. A. Rado). Convention date (U.S.A.) March 30th, 1946.

627 092.—Waveguide in which the tuning cup is supported by a dielectric ring bearing against a conducting surface of the guide.

Hazeltine Corporation (assignees of J. A. Rado). Convention date (U.S.A.) March 27th, 1946.

RECEIVING CIRCUITS AND APPARATUS

(See also under Television)

627 238.—Stability is ensured in a high gain i.f. amplifier by a frequency multiplier in the i.f. channel.

Bendix Aviation Corporation. Convention date (U.S.A.) December 13th, 1945.

627 412.—Tuning aid for f.m. receivers in which stray a.f. voltages, applied to the tuning control by the fingers, are used for amplitude modulation of the signal whereby circuit tuning is indicated by minimum response to such modulation.

Philco Corporation. Convention date (U.S.A.) February 26th, 1946.

627 985.—Amplitude-limiting detector circuit in which positive and negative half cycles are separately limited and combined with modulated oscillations to give a degree of asymmetry dependent on the original modulation.

Marconi's Wireless Telegraph Co. Ltd. (assignees of M. G. Crosby). Convention date (U.S.A.) July 31st, 1945.

TELEVISION CIRCUITS AND APPARATUS

FOR TRANSMISSION AND RECEPTION

627 528.—Scanning-beam transmitting tube in which all but the fastest of the repelled electrons are prevented from reaching the amplifying system.

Compagnie Pour La Fabrication des Compteurs et Materiel d'Usines A Gaz. Convention date (France) September 10th, 1946.

627 568.—Television receiver in which the tube is in a pivoted carrier so that the viewing screen can be raised above the top of the cabinet for use.

P.R.T. Laboratories Ltd., H. Bayhut and A. J. Gale. Application date June 5th, 1946.

627 588.—Saw-tooth generator operative only after receipt of a control pulse which is applied to a charging valve and to a valve controlling the discharging valve.

Sadir-Carpentier. Convention date (France) May 7th, 1945.

628 015.—Television system using time-modulated sound pulses superimposed on blanking pulses, the particular method of producing the composite signal being described.

Standard Telephones and Cables Ltd. (assignees of N. H. Young Jr.). Convention date (U.S.A.) March 30th, 1946.

CONSTRUCTION OF ELECTRON-DISCHARGE DEVICES

626 289.—A cavity-resonator valve in which electrons are injected from outside the cavity wall through a grid and having a reflector electrode within an opening opposite the grid, the said electrode being insulated so that it may have a different steady potential.

The British Thomson-Houston Company Limited, and W. J. Scott. Application date February 2nd, 1944.

626 687.—Combined thermionic valves and waveguides for high frequencies, in which one of the waveguides is provided with two spaced shorting bridges selected in position so that the impedance of the waveguide is matched to the valve.

The General Electric Company and E. G. James. Application date February 9th, 1944.

627 287.—Magnetron anode structure formed by the brazed assembly of metal rings plated with a fusible metal.

Marconi's Wireless Telegraph Co. Ltd. (assignees of L. P. Garner). Convention date (U.S.A.) May 30th, 1942.

627 980.—Velocity-modulation valve in which the end walls of the cavities are arranged as a deformable diaphragm for tuning control.

Compagnie Generale de Telegraphie sans Fil. Convention date (France) March 25th, 1944.

SUBSIDIARY APPARATUS AND MATERIALS

626 973.—A u.h.f. wavemeter embodies a resonant line tunable by a cup-shaped element adjustable around the end of an inner conductor of the line, means being included for indicating resonance.

W. A. W. Williams and B. Solley. Application date October 5th, 1946.

627 113.—A capacitor includes a compressed powdered ceramic dielectric whereby the capacitance can be varied by the voltage applied to the capacitor plates.

Radio Corporation of America (assignees of H. L. Donley and C. Wentworth). Convention date (U.S.A.) April 27th, 1946.

627 608.—Piezo-crystal circuit using several crystals of the same frequency with loss couplings between them to damp out unwanted frequencies.

Marconi's Wireless Telegraph Company Ltd. and W. S. Mortley. Application date December 5th, 1946.

627 870.—A band-pass filter of the resonant line type having tandem-connected parallel-disposed elements and coupling members arranged to disturb the natural coupling between said elements.

Hazeltine Corporation (assignees of H. A. Wheeler). Convention date (U.S.A.), November 8th, 1945.

www.americanradiohistory.com